
Sources of Predictability for Deep Convection: Radar Data Assimilation, Orography and Weather Regime

Kevin Anton Heinrich Bachmann



München 2019

Sources of Predictability for Deep Convection: Radar Data Assimilation, Orography and Weather Regime

Kevin Anton Heinrich Bachmann

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Kevin Anton Heinrich Bachmann
aus München

München, den 1.4.2019

Erstgutachter: Prof. Dr. George Craig
Zweitgutachter: Prof. Dr. Bernhard Mayer
Tag der mündlichen Prüfung: 13.5.2019

Machines take me by surprise with great frequency.
- Alan Turing

Parts of this thesis are published in:

Bachmann, K., Keil, C. and Weissmann, M.

Impact of radar data assimilation and orography on predictability of deep convection.
Q. J. R. Meteorol. Soc. 145, 117130 (2019). doi:10.1002/qj.3412

Bachmann, K., Keil, C., Craig, G. C., Weissmann, M. and Welzbacher, C.

Predictability of Deep Convection in Idealized and Operational Forecasts:
Effects of Radar Data Assimilation, Orography and Synoptic Weather Regime
Mon. Weather Rev. (2019)

Contents

Zusammenfassung	x
Abstract	xii
1 Introduction	1
1.1 Historic Developments	1
1.2 Review of the State of the Art	9
1.3 Arising Research Questions	13
2 Methods	15
2.1 Research Strategy	15
2.2 Forecast Model: COSMO	17
2.3 Data Assimilation Algorithm: KENDA	18
2.4 Experimental Setup	22
2.4.1 Idealized Model Setup	22
2.4.2 Operational Model Setup	28
2.5 Metrics and Scores	32
2.5.1 Convective Adjustment Timescale	32
2.5.2 Normalized Root Mean Square Error	33
2.5.3 Decorrelation Scale	33
2.5.4 FSS and Believable Scale	35
2.5.5 Comparison of the Predictability Metrics	37
3 Predictability of Convective Precipitation	41
3.1 Impact of Orography in Idealized Experiments	42
3.1.1 Development of Convection in the Nature Run	42
3.1.2 Orography and Initial Conditions in the Ensemble Forecasts	43
3.2 Impact of Radar Data Assimilation in Idealized Experiments	55
3.2.1 Orography and Initial Conditions under Radar DA	55
3.2.2 Sensitivities of the DA system	62
3.3 Impact of the Convective Mode in Idealized Experiments	66
3.4 Predictability of Convection in Operational NWP	72
3.4.1 COSMO KENDA: Impact of Orography and NWP system	72
3.4.2 COSMO-DE-EPS: Impact of Orography and Weather Regime	76

4	Conclusions	81
4.1	Orography and Weather Regime	83
4.2	Radar DA and Orography	85
4.3	Outlook on Future Applications	88
A	List of Abbreviations	91
B	List of Experiments	92
C	List of Figures	93
	Bibliography	99
	Acknowledgements	113
	Quotes	115

Zusammenfassung

Die Einführung von Konvektion auflösenden Vorhersagemodellen in den meisten operationellen Wetterzentren während der letzten 15 Jahre markiert einen Fortschritt in der Vorhersagbarkeit von Konvektion. Ein öffentliches Interesse an akkuraten Gewittervorhersagen ebenso wie die mit Gewittern verbundenen Sicherheitsrisiken machen diese zu einer zentrale Aufgabe für Wetterdienste. Eine Herausforderung ergibt sich hierbei aus der natürlichen Begrenztheit der Vorhersage auf wenige Stunden aufgrund der chaotischen Eigenschaften der Atmosphäre. Jedoch gibt es Hypothesen zu möglichen Quellen von Vorhersagbarkeit, die den Vorhersagehorizont auf den konvektiven Skalen erweitern können. Diese Skalen könnten Vorhersagbarkeit von anderen Merkmalen, wie z.B. Orographie oder dem Wetterregime, übernehmen. Darüberhinaus plant der Deutsche Wetterdienst noch 2019 operationell Radardaten zu assimilieren, was ebenso wie Orographie die Position von Konvektion beeinflusst. Wir nähern uns der Ermittlung der Grenzen von Gewittervorhersagbarkeit mittels einer Reihe von Modellkonfigurationen, kombiniert mit Radardatenassimilation und zweierlei Arten von Anfangsbedingungen, an. Wir verringern die Komplexität in einem idealisierten Setup mit vereinfachten Anfangsbedingungen (AB) und Orographie, ehe wir komplexe Orographie und die natürlich Variabilität des synoptischen Wetterregimes erneut einführen. Wir wenden eine Kombination von Verifikationsmethoden an und berücksichtigen so spezifische Facetten der praktischen Vorhersagbarkeit, d.h. der Modellvorhersagbarkeit und derjenigen der Atmosphäre. Im idealisierten Setup finden wir gesteigerte Vorhersagbarkeit von Konvektion unter Einfluss von Orographie, sofern die AB nur kleinskalige Fehler aufweisen, die eine perfekte synoptische Vorhersage repräsentieren. Wenn synoptische Fehler auftreten, wird die positive Wirkung der Orographie verringert. Jedoch kann die Assimilation von Radardaten diese Fehler kompensieren und hochwertige Analysen liefern. Der Vorhersagehorizont wird unter beiderlei AB um 6 Stunden erweitert, wobei Skalen bis zu 50 km vorhersagbar bleiben, wenn nur kleinskalige Fehler vorliegen. Wenn synoptische Unsicherheiten auftreten, sind Skalen bis 100 km vorhersagbar. Die Experimente zeigen ferner die gestiegene Vorhersagbarkeit bei Konvektion mit hohem Organisationsgrad. Auch konnten wir diese Ergebnisse im ab Mai 2019 operationellen COSMO-KENDA-System des DWD bestätigen. Des weiteren fanden wir eine höhere Vorhersagbarkeit von Regimen mit synoptischem Einfluss sowie im bergigen Süden Deutschlands verglichen mit dem flacheren Norden. Über drei Sommer hinweg fanden wir in einem älteren DWD-Modell, das Wetterlagen, die mehr von lokalen Prozessen abhängen und weniger vorhersagbar sind als diejenigen, die einem synoptischen Einfluss unterliegen. Beide Wetterlagen zeigen keinen signifikanten Effekt der Orographie.

Abstract

The advent of convection-permitting ensemble prediction systems at most operational weather centers within the last 15 years constitutes a step-change in our ability to forecast convection. This is a fundamental task for weather services, as not only the general public demands good and reliable forecasts of convection, but they can also be accompanied by heavy precipitation and destructive hail, and so comprise a risk to human life. Unfortunately, their prediction is challenging and their predictability limited by the chaotic nature of the atmosphere, so there is an intrinsic limit to their prediction in the order of a few hours. But for all that, the possible existence of sources of predictability that are able to extend the forecast horizon on the convective scales is being discussed. More precisely, those scales might inherit predictability from larger-scale features, such as orography or the prevailing weather regime. Furthermore, radar data will be operationally assimilated at Deutscher Wetterdienst (DWD) within the year and provide similar information as the natural sources of predictability—the position of convection. The identification of predictability limits of convection poses significant challenges, which we address within a hierarchy of model configurations, combined with radar DA and two sets of initial conditions (ICs). First, we reduce the complexity in an idealized setup with simplified ICs and orography before we reintroduce complex orography and natural variability of the synoptic weather regime. We apply a combination of sophisticated verification metrics to address specific facets of practical predictability, namely the predictability of the model state and the model predictability of the atmospheric state. In the idealized setup, we find increased predictability of convection in the presence of orography if the ICs depict only small-scale errors, representing perfect large-scale predictability. In the presence of large-scale errors, the beneficial effect of the orography is greatly diminished. However, the assimilation of radar observations proved its ability to account for these errors and provide high-quality analyses. In both sets of ICs, the forecast horizon is extended by 6 h, where scales up to 50 km remain predictable for small-scale errors, and scales up to 100 km do so with additional large-scale uncertainty. The experiments also highlight increased predictability for convection with a high level of organization. We were also able to confirm these results in the pre-operational COSMO-KENDA system of the DWD. More specifically, we found increased predictability of the model state and model predictability of the atmospheric state in the orographically more influenced South than in the comparatively plain North of Germany. Throughout three summers in an older DWD system, we found locally forced weather situations to be less predictable than those forced by the synoptic weather regime and no significant effect of orography.

Introduction

1.1. Historic Developments

Towards operational ensemble prediction

Numerical Weather Prediction dates back more than 100 years, when Vilhelm Bjerknes first proposed the concept of a mathematical model of the atmosphere's dynamics (Bjerknes, 1904). Weather forecasting then becomes a problem with two parts: First, the determination of the initial state of the atmosphere, and second, the solution of the governing equations of the system. Lewis Fry Richardson was the first to attempt a real numerical weather forecast by manually solving the primitive equations, which took him about six weeks (Richardson, 1922). Although his six-hour forecast of the surface pressure over two points in Central Europe yielded unrealistic results due to imbalances in the initial conditions (Lynch, 2008), he permanently established the basic method of modern-day Numerical Weather Prediction (NWP).

It took another 25 years until a successful forecast was performed. John von Neumann recognized weather prediction as a perfect application for the electronic computer he and his team had developed at the Institute for Advanced Studies in Princeton between 1946 and 1952. Therefore, one out of the four project groups was 'Meteorology', directed by Jule Charney between 1948 and 1956. By 1950, Charney and his group had simplified the primitive to the quasi-geostrophic equations, developed a numerical algorithm and tested it at the Electronic Numerical Integrator and Computer (ENIAC) in Aberdeen, Maryland (Charney et al., 1950). The 24-hour forecasts each took almost 24 hours to compute, but they indeed resembled the large-scale features of the mid-tropospheric flow. Only five years later, their success would lead to the first operational NWP in the United States in 1955 (Lynch, 2008).

As already mentioned, calculating the governing equations is only part of the problem, and the failure of Richardson's forecast demonstrates the crucial importance of determining the atmosphere's initial state, which is not a straightforward problem and

resulted in the emergence of Data Assimilation (DA) as a central component of NWP. DA combines the sparse, erroneous and sometimes indirect observations of the atmosphere with a physically consistent model state of the atmosphere that includes all prognostic variables, referred to as the analysis. Early attempts were based on simple interpolations, which were soon superseded by so-called statistical interpolation methods (Gandin, 1966), which account for errors in the NWP model and the observations. DA methods advanced over the decades, and today sophisticated algorithms, like Four Dimensional Variational Data Assimilation (4DVar) and the Ensemble Kalman Filter (EnKF) are applied (Evensen, 1994, 2003).

Bauer et al. (2015) acknowledge the advances made in the quality of forecasts since then as 'the quiet revolution of numerical weather prediction', based on a steady improvement of model and observation systems in combination with increasing computing power, described by Moore's Law. Over the last 40 years, the forecast skill of medium-range forecasts increased by about one day per decade, which means that today's 7-day forecast is as good as the 6-day forecast ten years ago. Also today, the skills in the Northern and Southern hemispheres are almost equal as the result of a denser global observation network, including the assimilation of satellite data.

Predictability of atmospheric flows

Nonetheless, forecasts are not always right, and the reasons are manifold. The most fundamental one is the intrinsically limited predictability of atmospheric motions. In 1961, Edward Lorenz found that he was unable to repeat a previous computation of an NWP model and tracked the problem to small truncation errors in the ICs. The errors mentioned above would double every four days, and after about two months, any resemblance to the original results was gone. His finding would lead to the foundation of chaos theory as a scientific field and seminal papers in atmospheric predictability research (Lorenz, 1963, 1969).

Lorenz examined the temporal evolution of small IC errors in a low-dimensional flow model with only three degrees of freedom, the so-called Lorenz attractor. Although the system is deterministic without any stochastic elements and ICs and model equations are known, the differences in a pair of simulations with perturbed ICs, usually called identical twins, would grow exponentially until they are indistinguishable to random flow realizations. This sensitivity of a deterministic system to small errors in the ICs is called deterministic chaos. Unfortunately, the real atmosphere is such a chaotic system with many more degrees of freedom. Moreover, the ICs are subject to measurement errors and representativity problems, and the governing physical equations cannot be solved analytically and have to be discretized.

Probabilistic forecasting

The realization of the existence of an inherent limit to the predictability of the atmosphere led to a shift in meteorology. The early works of Abbe (1901) and Bjerknes

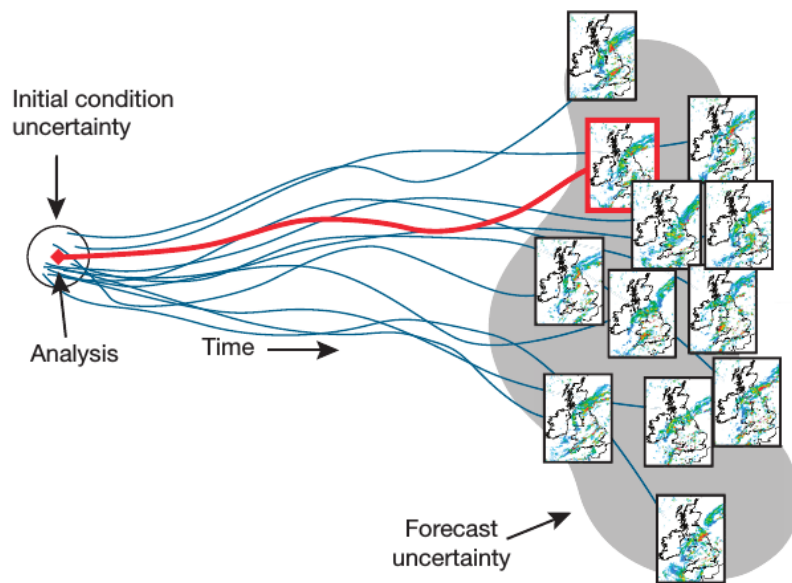


Figure 1.1: Schematic diagram visualizing the use of an ensemble to estimate forecast uncertainty. In contrast to a single deterministic forecast (red frame), an ensemble (black frames) with forecast uncertainty (grey shading) is computed. The figure is taken from Bauer et al. (2015).

(1904) show a strong belief in the possibility of deterministic long-range weather prediction in the sense of Laplace's demon (Laplace, 2012). In contrast, the works of Thompson (1957) and Lorenz (1963) point towards an alternative probabilistic approach to NWP, where the IC uncertainty is sampled with a Monte-Carlo method, and a so-called ensemble of states is propagated in time. It was first put in operation at the European Centre for Medium-Range Weather Forecasts (ECMWF) in the early 1990s, when the computational demands of an ensemble approach could be met and became an established procedure at NWP centers worldwide to deal with inherent uncertainties in the predictions (Leutbecher and Palmer, 2008).

A schematic visualization of the ensemble approach is depicted in Fig. 1.1. The red line illustrates a single deterministic forecast propagated in time and resulting in a precipitation forecast for Great Britain. In contrast, an ensemble accounts for the IC and model uncertainty in the form of an analysis, which is provided by a sophisticated DA system. In practice, the uncertainty is not sampled randomly (Monte-Carlo), but in a way that guarantees sufficient error growth. The initially small uncertainty grows over time in a non-linear fashion (blue lines), resulting in an ensemble forecast with different outcomes that can be used to produce probabilistic forecasts.

Intrinsic and practical predictability

At this point, it is useful to distinguish two fundamental concepts—intrinsic and practical predictability. Intrinsic predictability, or the Lorenz limit, refers to the extent to

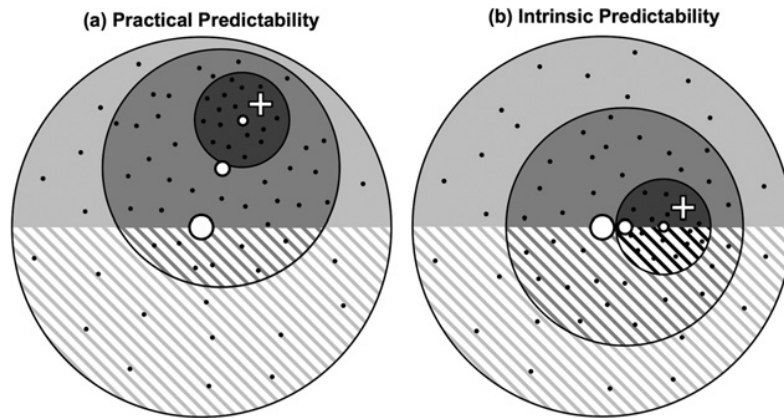


Figure 1.2: Idealized schematic illustrating the difference between practical (a) and intrinsic (b) predictability. The circles represent ensemble forecast with the ensemble members as black dots; white dots represent the respective ensemble mean and the white cross the truth. Shadings indicate two different flow regimes. The figure is taken from Melhauser and Zhang (2012).

which prediction is possible if an optimum procedure is used in the presence of infinitesimal IC errors (Lorenz, 1969). The intrinsic limit is insurmountable due to the chaotic nature of the atmosphere, and most notably due to moist convective processes (Zhang et al., 2003; Hohenegger et al., 2006; Melhauser and Zhang, 2012). In contrast, practical predictability describes the ability to predict based on the procedures currently available. In reality, errors in the ICs and the forecasting system limit this ability (Lorenz, 1996). Just like the intrinsic predictability, the practical limits of the models we use to predict the atmosphere impose limits to NWP. Therefore, practical predictability limits of atmospheric models have to be quantified.

Figure 1.2 schematically illustrates the differences between intrinsic and practical predictability. The circles represent the IC uncertainty with the ensemble members (black dots) and the ensemble mean in the center (white dot). The different shadings indicate two different possible outcomes, for example, the development of a severe thunderstorm and a harmless cumulus cloud, with the white cross indicating the actual outcome. An improvement of the forecasting system or of the observations results in a reduced IC uncertainty (smaller circle) and, in a system limited by practical predictability, an improved forecast. Consequently, all ensemble members in Fig. 1.2a agree on the right general outcome, although the forecast might still be flawed in some other sense. In the limit of intrinsic predictability, the two outcomes are equally likely, and the atmosphere is close to a transition point that means prone to IC uncertainty. In this case, the same reduction in the IC errors does not result in a forecast improvement. The ensemble in Fig. 1.2b is, even after a substantial decrease of the IC uncertainty, unable to predict the truth with certainty as it still incorporates both possible outcomes.

This dissertation concentrates on the practical predictability but makes an additional distinction. Similar to Surcel et al. (2015), we distinguish further between the predictability of the model state and the model predictability of the atmospheric state. The predictability of the model state describes how a process or effect influences the

development of the NWP system itself. The model is used as a proxy for the atmosphere, and to the degree of which this is true, the predictability of the model state is a proxy for the intrinsic predictability of the atmosphere. It depends on the forecast model, the DA and the verification metric. The model predictability of the atmospheric state additionally incorporates a comparison to observations. The latter therefore includes model and observing system errors and represents a classic practical predictability estimate. Examples for significant inaccuracies in NWP models are the discretization or numerical representation of the governing equations and physical processes unresolved by this discretization, like radiation, turbulence, and microphysics, that have to be parameterized. Errors in the observing systems are linked to measurement errors, as well as quality control and spatial and temporal coverage (Sun and Zhang, 2016).

Flow-dependent predictability

Figure 1.3 shows the Lorenz attractor and visualizes another critical feature of chaotic systems and the atmosphere—predictability is flow-dependent. The yellow shading indicates the climatology of the Lorenz attractor in two dimensions and the black circle the initial uncertainty of an ensemble forecast. Arrows and resulting circles express the ensemble evolution with time. Fig. 1.3a displays a highly predictable case, which can be seen in the coherence of the ensemble in the forecast. The ensemble evolves through a stable part of the Lorenz attractor, resulting in a sharp probabilistic forecast. Figure 1.3c, in contrast, exhibits possible forecast outcomes in large portions of the model climatology and therefore low predictability as it is initialized close to a particularly unstable part of the attractor. Figure 1.3b is an intermediate case between those extremes. The flow-dependent predictability of a chaotic system has direct consequences in meteorology. Some weather situations or phenomena may be highly predictable, while others show almost none (Zhang et al., 2006). This implies that the predictability is variable and has to be predicted itself. The information about the uncertainty of a forecast is viable and as Tim Palmer put it: ‘no weather or climate prediction can be considered complete without a forecast of the associated flow-dependent predictability’ (Palmer, 2000).

Scale-dependent predictability

Lorenz (1969) also considered scale interactions of a ‘flow which possesses many scales of motion’, very much like the atmosphere, as weather phenomena span many orders of magnitude from planetary waves with scales of $\mathcal{O}(10\,000\text{ km})$ over single clouds of a few hundred meters down to turbulence on the very smallest scales. Figure 1.4 shows a reproduction of a key figure from his paper illustrating the scale dependence of predictability. The top curve represents an assumed background energy spectrum of the atmosphere as a function of the horizontal scale. Nastrom and Gage (1985) later show, based on aircraft wind measurements, that the underlying properties of the spectrum, like larger energy on larger scales, hold in reality, although the spectral slope is unrealistic. The intersecting lines show scale and amplitude of energy of a forecast

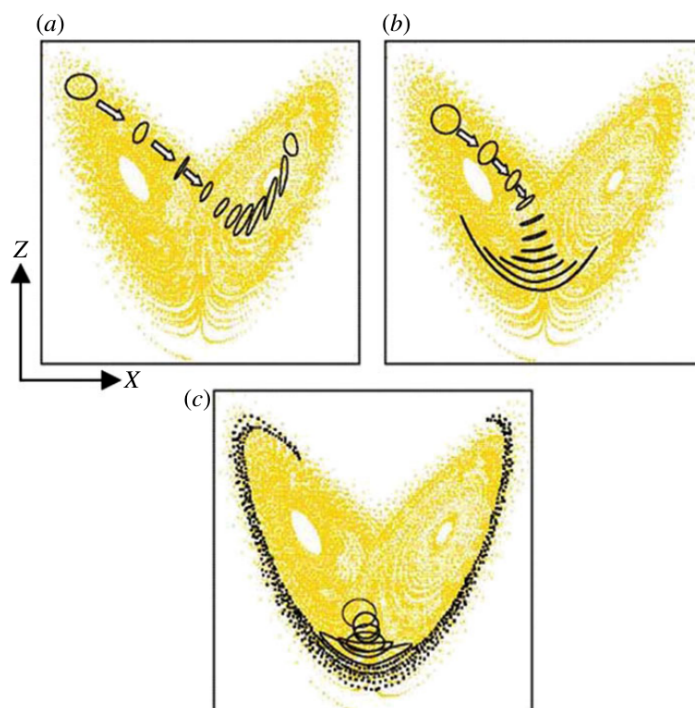


Figure 1.3: Phase-space evolution on the Lorenz attractor for three probabilistic predictions, which are started from different points on the attractor highlighting the flow dependence of the predictability. Examples of finite-time error growth in a highly predictable case (a), a case with high initial predictability but increasing uncertainty in the course of the forecast (b), and a forecast starting near the transition point between regimes exhibiting low predictability (c). The figure is taken from Slingo and Palmer (2011).

error that was initially constrained to large scales. In his experiments, those large-scale errors would project on the smallest resolved scales almost immediately and propagate upscale from there. The error energy saturates when it grows to the same amplitude as the background energy. This happens on ever-growing scales causing a continuous loss of predictability from small to large scales. The estimated forecast horizon, or the time period in which a skillful forecast is possible, of a thunderstorm with a horizontal scale of around 40 km is 1 h, while large-scale wave patterns like Rossby waves remain predictable for up to 5 days. The described behavior is referred to as the *butterfly effect* as Lorenz understood it.

Importance of convective-scale forecasting

Notwithstanding the apparent difficulties, forecasting thunderstorms is an essential issue in operational weather prediction as they are often accompanied by heavy precipitation, gusts, and destructive hail, which have considerable social and economic consequences in the United States and Central Europe (Ashley and Ashley, 2008). A recent study reports that the risk of a fatal car crash increases by 34 % during heavy precipitation, which caused about 10,000 fatalities in the United States within the investigated six year period Stevens et al. (2019). Another example is the High Impact Weather (HIW)

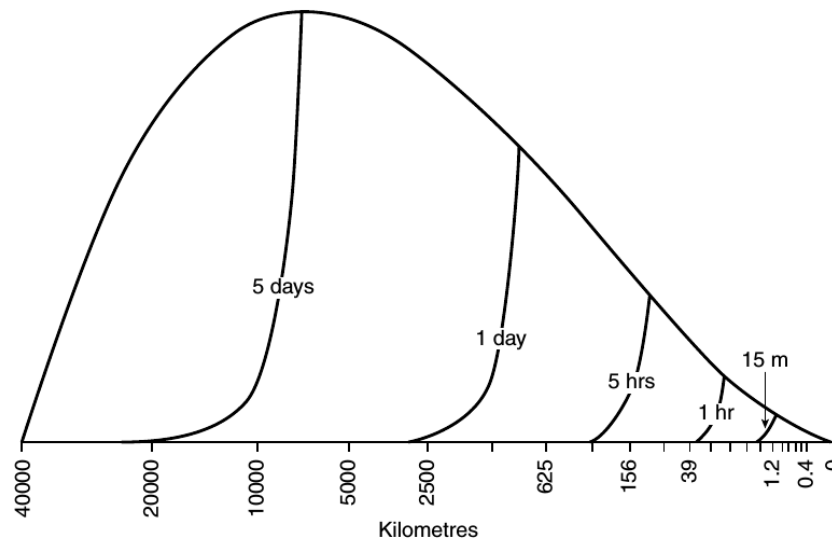


Figure 1.4: Scale-dependent predictability limits from Lorenz (1969): The top curve shows the background or saturation energy as a function of horizontal scale. Areas are proportional to the energy. The annotated lines depict the error evolution with time and coincide with the top curve to their right. The figure is taken from Lorenz (1969) and reproduced by Palmer et al. (2014).

period in May and June 2016 in Central Europe (Piper et al., 2016). Additionally, several studies in recent years linked misrepresented convection over North America to midrange forecast busts, periods of especially bad skill, over Europe (Rodwell et al., 2013). As a result, an ongoing discussion developed if improved representation and prediction of convection could improve synoptic-scale forecasts (Durran and Gingrich, 2014; Durran and Weyn, 2016; Sun and Zhang, 2016).

The abovementioned estimates for the storm-scale forecast horizon might be too pessimistic, but more recent studies only extend the estimates to a few hours (Hohenegger and Schär, 2007a,b; Zhang et al., 2015, 2016; Judt, 2018; Zhang et al., 2019). So far, we have established some of the inherent difficulties every study on the predictability of convective precipitation, just like the present one, has to face. Deep convection is located in the realms of low predictability as it is a small-scale feature in a particularly unpredictable flow regime. Various publications have linked the latent heat release in convective clouds to rapid error growth, which indicates that convection itself can constitute a transition point in the forecast (Zhang et al., 2003, 2007; Selz and Craig, 2015; Zhang et al., 2016).

Sources of atmospheric predictability

Anthes (1986) hypothesized that so-called ‘sources of predictability’ could enhance the forecast horizon on the convective scale behind the proposed limits and mentioned orography or synoptic-scale weather regimes as possible examples. More recently Bauer et al. (2015) and Yano et al. (2018) referred to orography and large-scale forcing of

convective-scale weather as sources of atmospheric predictability. The reasoning behind this is that NWP models are in general able to predict the meteorological situation, like temperature or humidity, on a scale of $\mathcal{O}(100\text{ km})$, but miss the correct small-scale variability that is necessary to predict the exact location where a convective cell is triggered and subsequently develops. Orography provides a potent trigger for convection by several mechanisms including the forced lifting of air masses, flow convergence in its lee, or differential heating on its slopes (Houze, 2012). Its relevance is shown by Carbone and Tuttle (2008) and Levizzani et al. (2010), who link over 50 % of summer precipitation in the Great Plains and the Mediterranean to elevated terrain.

Similarly, synoptic-scale weather causes features like convergence lines or fronts that can trigger low-level convergence in confined regions, which is a major contributor to convective initiation. Therefore, those sources provide potent triggers that influence the probability of convective triggering. The consecutive evolution of convection is not necessarily affected, but can trigger further convection, thereby transporting predictability in space and time.

Requirements for investigation of predictability sources

Influenced by Anthes (1986) and the increasing computational capabilities, Lilly (1990) raised the challenging question of whether the time of NWP of thunderstorms and on convective scales has come. He identified the critical challenges of convective-scale NWP that should be addressed in the following years. As the model resolution of a few kilometers necessary to simulate convection explicitly is computationally demanding, Limited-Area Models (LAMs), with sophisticated physics parametrizations are needed. Those models need to be initialized from analysis states based on spatially and temporally dense observations like Doppler radar reflectivity. Only later was the considerable challenge of verifying the high-resolution forecasts of such systems with observations was acknowledged (Ahijevych et al., 2009; Cintineo and Stensrud, 2013; Clark et al., 2016).

Douglas Lilly underestimated the scientific and technical challenges that are linked to forecasting convective-scale weather and especially precipitation. More than 25 years later, those challenges have been addressed to a degree that makes operational weather forecasts of summer convection feasible (Clark et al., 2016) and, consequently, operational weather centers move towards convection-permitting models in combination with high-resolution radar observations and, for the first time, cycling windows of three hours or less (Gustafsson et al., 2017). In the light of these developments, it is seminal to investigate the forecast horizon of convective predictability in these NWP systems as a function of time and horizontal scale, as well as their dependence on orography, the prevailing weather regime or the combined impact of both. The advent of those sophisticated weather models also allows us to address questions regarding the effects of radar DA in these configurations. The answers to those questions will help forecasters to provide more reliable guidance and communicate their work and the associated uncertainties more effectively. Operational centers might identify the results as advantageous to allocate resources in the future.

1.2. Review of the State of the Art

In the following, we want to provide the reader with an overview of studies in the research fields that are vital for this dissertation. Therefore, we will briefly summarize the development of convection-permitting Ensemble Prediction Systems (EPSs), the application of EnKFs on convection-permitting resolutions and the issues that arise from verification of precipitation forecasts on these high resolutions, as well as approaches to handle them. Subsequently, we will present the subjects of convective-scale predictability, interactions of orography with the atmosphere and the impact of the prevailing synoptic weather regimes on convection.

Numerical modeling on convective scales

We refer to convection-permitting models, as those with a horizontal grid spacing between 1 to 5 km, which resolve convection explicitly and without parametrization (Nielsen and Schumacher, 2016). Those models resolve convective systems, but not the details of single convective cells like their updrafts, as a *convection-resolving* model would. However, many studies have shown the value of convection-permitting models as a forecast and research tool (Gebhardt et al., 2008; Schwartz et al., 2010) and their provided value over coarser resolutions, especially in convectively active regimes (Kain et al., 2013).

Mass et al. (2002) and Bryan et al. (2003) found generally improved realism of convection-permitting forecasts, but those do not necessarily lead to an improvement in objective scores. Similarly, Done et al. (2004) report identifiable Mesoscale Convective Systems (MCSs) and improved realism in 4 km resolution convection-permitting forecasts compared to 10 km resolution forecasts with parameterized convection for a two months summer period. Although the location-specific prediction is not enhanced, the convective organization type and the number of convective cells is. During the 2007 National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) Spring Experiment, the convection-permitting model forecasts proved not only to be more realistic, but also showed higher skill over 36 hours compared to 12 km simulations (Schwartz et al., 2009). All in all, precipitation forecasts of convection-permitting models still tend to be under-dispersive or, in other words, over-confident (Romine et al., 2014; Dey et al., 2016).

Data assimilation on convective scales

Convection-permitting NWP also calls for the assimilation of temporally and spatially highly resolved observations like radar data. Early studies explored the utility of the EnKF for this purpose and found promising results, while also struggling with model

errors producing large forecast errors after only 60 min (Snyder and Zhang, 2003; Zhang et al., 2004). As a consequence, the first operational applications of radar DA were based on latent heat nudging (LHN) due to higher consistency and simpler implementation (for example Stephan et al., 2008; Dixon et al., 2009, at DWD and MetOffice). However, the ongoing interest in the EnKF has been proven by the vast amount of case studies with various focus, like assimilation of different convective modes (Aksoy et al., 2009, 2010), continuous improvements in the representation of one specific extreme storm in Oklahoma (Snook et al., 2011, 2012, 2015; Putnam et al., 2017) or a severe weather outbreak in Alabama (Yussouf et al., 2015). Similarly, Potvin and Wicker (2013) investigate the capabilities of the NOAA Warn-on-Forecast system in an Observing System Simulation Experiment (OSSE) of a tornado. More OSSE studies address effects of the higher scanning frequency the phased-array radar (PAR) technology provides (Yussouf and Stensrud, 2010; Supinie et al., 2017), as well as the synergies that might arise from the combination of reflectivities measured by radar and brightness temperatures measured by satellite (Wheatley et al., 2015; Jones et al., 2016; Cintineo et al., 2016) and the choice of localization within the DA algorithm (Sobash and Stensrud, 2013; Lange and Craig, 2014).

All these studies mention improved analysis states after assimilation of radar DA, while the resulting forecasts degrade in skill rapidly due to model errors and physical instabilities that result in investigated lead times of only up to 90 min (up to 3 h in Putnam et al., 2017). They also focus on the study of just a few convective events, which are often extreme and close to certain densely observed areas, like Oklahoma, which limits their representativity. More recently, however, Johnson et al. (2015) found a beneficial impact of radar DA with an EnKF for up to 5 h and better skill than with Three Dimensional Variational Data Assimilation (3DVar), which can be attributed to the lack of flow-dependence and cross-variable correlations in this system. First experiments with full DA-model cycling on convective scales followed and highlighted the vast potential of the EnKF for operational convective-scale applications in a two months period in May and June 2013 in the United States (Schwartz et al., 2015), a period of seven consecutive days in Germany (Bick et al., 2016) and four days in Italy (Gastaldo et al., 2018). In conclusion, the developments in recent years regarding cloud-permitting models combined with radar DA based on an EnKF provide forecasts of convection, that are not only realistic, but also skillful for an extended range of several hours.

Verification of convective-scale forecasts

Simultaneously to the model developments, new verification methods were introduced to and adopted by the community. Forecasts of convection with their inherently high spatio-temporal variability and intrinsically low predictability require spatial metrics, also called neighborhood or fuzzy metrics, as they are prone to double penalties (Casati et al., 2004; Gilleland et al., 2009). Traditional scores, like the Root Mean Squared Error (RMSE), favor smoother forecasts and penalize the details at a higher resolution twice, unless the precipitation is predicted in the exactly correct position because a mismatch with the truth gives both a miss and a false alarm. Typically, a forecaster will find such a slightly displaced forecast useful nonetheless. One of the more popular approaches

to deal with double penalties is the Fractions Skill Score (FSS) introduced by Roberts and Lean (2008) and its extension to the believable scale (Mittermaier et al., 2013). The believable scale provides, similar to the decorrelation scale (Surcel et al., 2014, 2015), an agreement scale of the ensemble forecasts that is easy to interpret for spatial structures such as precipitation. Weyn and Durran (2018a) recognized the seminal importance of neighborhood verification, as only the FSS was able to differentiate convective predictability in ensemble forecasts, while the classical error growth failed to depict differences. This is probably due to the more local and immediate consequence of precipitation compared to the smoother and larger wind fields. In addition, object-based verification like the Structure, Amplitude and Location Score (SAL) (Wernli et al., 2008) or the Displacement and Amplitude Score (DAS) (Keil and Craig, 2009) were introduced.

Convective scale predictability limits

The facts that the predictability of the atmosphere is intrinsically limited, and that the forecast horizon is shorter on convective than on synoptic scales, are undisputed; however, quantifying the limits for convection in different situations has been an active topic in the last decade. The approaches to do so are varied and range from conceptual arguments over idealized numerical studies to case studies. Rotunno and Snyder (2008) and Durran and Gingrich (2014) determined the limits of predictability based on error growth and the error kinetic energy saturation in a generalized Lorenz model. Both describe a crucial dependence on the background spectrum of the atmosphere, which ranged from unlimited predictability in a spectrum with slope ' k^{-3} ' to limited predictability for a ' $k^{-5/3}$ ' spectrum, which is more realistic for the mesoscales. In a more applied publication, Hohenegger et al. (2006) found that the predictability of convection in three cases of the Mesoscale Alpine Programme (MAP) is highly case-dependent, and differences in predictability are not obvious to explain, followed by the finding that the tangent-linear approximation fails on cloud-resolving scales after only $\mathcal{O}(1.5\text{ h})$ (Hohenegger and Schär, 2007b), which renders prediction after this point inherently difficult as non-linear processes become dominant.

Zhang et al. (2007) link the rapid error growth in a case study to moist convection and latent heat release and explain very short forecast horizons with model errors and poor analysis. A similar survey by Melhauser and Zhang (2012) reports improvements of the model, but raises the question of whether the predictability of the specific case, a bow echo event in North America, might be intrinsically limited. Further studies focus on a comparison of intrinsic and practical limits in the case of a tornadic thunderstorm near Oklahoma on 20 May 2013 (Zhang et al., 2015, 2016). Interestingly, a sensitivity of the predictability to the orography in the region was found, as it constrains the low-level convergence to the correct area. Along with Barrett et al. (2015), who report increased predictability of a stationary rain band over the United Kingdom due to interaction with orography, those are the only recent studies on predictability connected to orography known to the author.

Two more publications have to be mentioned in the context of predictability of precipitation in the presence of orography. Walser et al. (2004) and Walser and Schär

(2004) investigated the predictability of convection in the Alpine region for four cases and focus on quantitative precipitation forecasts for river catchments. Their results suggest that while local predictability is lost rapidly due to chaotic aspects of the moist dynamics, the prediction for a larger area like a catchment can be skillful, although the skill is highly case-dependent.

Similar to the publications regarding the assimilation of radar observations, the research is mainly limited to a few often extreme case studies or idealized experiments initialized with so-called warm bubbles (Snyder and Zhang, 2003; Zhang et al., 2004; Sun et al., 2017; Weyn and Durran, 2017). However, warm bubbles are suspected of disrupting the model dynamics unnaturally and impacting subsequent forecasts.

Apart from the presented publications, several authors published multi-annual climatologies of observed precipitation, which depict certain patterns of maxima and minima that can be linked to orography, availability of surface humidity or typical weather situations in the region. Examples are Kuo and Orville (1973) and Kovacs and Kirshbaum (2016) for the Black Hills and Southern Quebec, Levizzani et al. (2010) for Europe and the Mediterranean, while Isotta et al. (2014) focused on the Alpine region and Foresti et al. (2018) on the Swiss Alps. Those climatologies hint towards modified predictability and their causes but are unable to quantify it. Moreover, correlation does not imply causality. More fundamental understanding of the processes is desirable, and various publications cover the interactions between orography and the atmosphere.

Predictability in the presence of orography

As mentioned, literature directed towards convective predictability is rare, but a profound amount of work has been done in other areas of orographic effects. Several studies describe the precise mechanisms with which orography can trigger convection, like, for example, thermal forcing on slopes (Demko and Geerts, 2010), channeling of airflow near mountains or diurnally triggered disturbances propagating away from them (Houze, 2012), while others focus on micro-physical impacts (Zwiebel et al., 2016; Panosetti et al., 2016). During the Convective and Orographically-induced Precipitation Study (COPS), which resulted in several papers based on a measurement campaign in the Black Forest in summer 2007, several convective cells were observed and investigated. The studies agree on a high sensitivity of the results on ICs and highlight the importance of boundary-layer convergence, as well as the interaction of various scales to trigger convection (Kirshbaum, 2011; Corsmeier et al., 2011; Barthlott et al., 2011).

Predictability in different weather regimes

Several methods to define synoptic-weather regimes have been advised, but one of particular usefulness for applications regarding summer precipitation is the convective adjustment timescale (Done et al., 2006). It describes a spectrum from situations with strong synoptic-scale forcing to those with weak forcing, which are more reliant on

local triggers for convective initiation. Studies that apply the convective adjustment timescale or comparable metrics tend to focus on two main points.

First, the sensitivity to IC uncertainties on small compared to large scales is investigated in both regimes. They show that the synoptically forced regime is more resilient towards small-scale uncertainties, which only exhibit a comparable effect as large-scale uncertainties in locally forced cases (Johnson et al., 2014; Flack et al., 2018; Weyn and Durran, 2018b). Those results are in line with the fact that the impact of radar DA by LHN is extended in synoptically forced situations (Craig et al., 2012). Secondly, a range of studies linked the weather regime directly to the predictability of convection in various regions of the world, like the Alps (Walser and Schär, 2004), Germany (Keil et al., 2014), the United Kingdom (Flack et al., 2016) and the United States (Surcel et al., 2017).

1.3. Arising Research Questions

The existence of sources of predictability for deep convection, like orography and the weather regime, are a long-standing hypothesis in meteorology. The increasing computing power and the resulting development of convection-permitting NWP allow us to revisit this proposition by means of numerical experiments. In addition, recent advancements in the assimilation of radar measurements provide the opportunity to investigate its impact compared to orography, which as a trigger for convection also provides information on the location of convection. Another source of predictability addressed in this thesis is the level of convective organization.

This dissertation is unique in the respect that it combines the abovementioned prerequisites that have only been available for a few years: a state-of-the-art convection-allowing NWP model with advanced convective-scale DA of radar measurements. Subsequently, the aspects of practical predictability are assessed with modern verification methods.

To conclude this introduction, we want to lay out our specific research questions in the hope that the overview of predictability studies based on convection-permitting EPS and their limitations, as well as the lack of research concerning orography and its apparent link to radar DA, have been able to convince the reader of their importance:

- How do sources of predictability, such as orography, influence the predictability of convective precipitation?
- What is the potential impact of radar data assimilation on the practical predictability of convective precipitation, and is its impact affected by orography?
- What is the impact of the convective mode or the level of organization on the predictability?

- In which way do sources of predictability, like orography or the synoptic weather regime, impact the practical predictability of deep convection in an operational NWP system over Germany these days?

The outline of this dissertation is as follows. Chapter 2 describes our research strategy and, based on it, the NWP system, consisting of a weather model and a DA algorithm, and the applied model configurations. It also introduces the metrics and scores used in this thesis to assess predictability and the prevailing weather regime. The results concerning the predictability of convection are presented in Chapter 3. It includes different setups of the NWP system to estimate impacts of radar DA, orography and weather regime. The findings are summarized and discussed in a broader context in Chapter 4.

Methods

In this chapter, the strategy to answer the posed questions and the required tools are introduced. The model components, the hierarchy of model configurations and the verification strategy are explained. This includes descriptions of the applied NWP model, DA algorithm and the range of applied configurations. Additionally, the convective adjustment time scale used to define the prevailing weather regime and the predictability metrics are described. As those metrics are especially crucial for this dissertation, their properties will be compared, and differences highlighted.

2.1. Research Strategy

The questions raised in the introduction ask for numerical experiments with certain capabilities. We require a convection-permitting NWP model that provides a realistic representation of convection, and as we are interested in practical predictability in Germany, the Consortium for small-scale modeling model (COSMO) developed at DWD is an obvious choice. It is designed for operational and scientific applications on convective scales and is being used operationally in varying configurations since April 2007, which also means it is well-tested and understood. Operationally, it is used as a LAM over Central Europe, but it can also be used in an idealized OSSE setup. Bierdel et al. (2016) showed that COSMO produces realistic kinetic energy spectra, which is seminal for error growth and predictability studies, such as the present.

Naturally, we coupled COSMO to the Kilometre-Scale Ensemble Data Assimilation (KENDA), which is an example for an EnKF and one of the DA algorithms developed and used at DWD. It provides various essential advantages. From a technical viewpoint, the coupling to COSMO was already established for research purposes at DWD, and we included the possibility to use it within the idealized OSSE setup. And more importantly, from an algorithmic viewpoint, KENDA is able to account for flow-dependent uncertainties derived from an ensemble. As pointed out, the predictability of the atmosphere is flow-dependent and, therefore, KENDA is, in contrast to a 3DVar algorithm with a climatological error covariance matrix, a perfect tool for this study. Additionally,

it was used successfully in previous studies to assimilate radar data into COSMO in both idealized (Lange, 2016) and pre-operational studies (Bick et al., 2016).

Nonetheless, the identification of predictability limits of deep convection poses a significant challenge using a state-of-the-art EPSs. The complexity of the problem lies both within the intermittent, spatio-temporally highly variable nature of convective precipitation depending on an ever-changing weather situation and within the different components of modern convective-scale NWP systems. Those include the challenge to assimilate high-quality observations to provide the initial state of the atmosphere, as well as an appropriate scale-dependent verification.

As a first step, we strongly simplify the problem and perform numerical experiments with the COSMO-KENDA system in an idealized OSSE. To aim at the impact of DA on the predictability limits of convective precipitation in the presence of orography, we exclude several influences: (i) the inherent weather situation uncertainty by prescribing horizontally homogeneous atmospheric conditions with periodic boundaries susceptible to long-lasting convection in Central Europe, (ii) observation uncertainties as representativity issues or systematic and correlated errors by using a perfect model OSSE approach creating spatio-temporally highly resolved *synthetic* observations from a Nature Run, (iii) thereby neglecting approximations and inconsistencies in the description of atmospheric processes in the NWP model or the forward operator, and (iv) the in reality complex orography is replaced with a simple Gaussian mountain preventing interactions of complex terrain with the atmospheric flow.

The initialization from a homogeneous background that is realistic and susceptible to long-lasting convection with subsequent spin-up provides various advantages. The ensemble forecasts are representative for a single weather regime, even if different convective modes develop within an ensemble. Additionally, as the convection is triggered by small amplitude noise perturbations in the boundary layer, its dynamics are not disrupted by any kind of artificial triggering like bubbles. The noise also provides no favored area for convective initiation, which represents an especially challenging situation for the DA algorithm. In this configuration, we can also test the impact of uncertainties on the largest resolved scales of the LAM, which are in reality inadvertently inherited from an ultimately global model, on the predictability limits of the forecasts. The idealized configuration allows us to control the amplitude and structure of those uncertainties.

Additionally, we can investigate specific facets of DA sensitivities. The idealized setup allows looking into the assimilation of a sub-sample of the observations, such as wind or reflectivity observations, which might be more realistic to accomplish for operational applications at many forecasting centers soon, or observation errors of our idealized radar and estimate effects of future technological improvements. We can also test if the convective type or the advancement in terms of the convective life cycle of the convective cells has an impact.

Every idealized study has its limitations, and the validity of the results for realistic applications can be unclear. Therefore, we perform further experiments with the same NWP system, but in a more realistic configuration. We forecast a two weeks HIW period over Germany that features a similar synoptic weather situation as the idealized study,

but with the natural variability of the real atmosphere. The German domain features complex orography and can be divided in a Northern and a Southern part, which fortunately possess a very different amount of orography. This allows us to attribute differences between the sub-domains to the complex terrain. This configuration is based on the pre-operational configuration that will become the regional operational model for Germany in May 2019, and facilitates comparisons to the LHN that has been in operation so far.

In a last step towards more realism, operational forecasts of the summer periods between 2014–2016 are analyzed. The large number of forecasts over this elongated and multi-annual period covers, in contrast to the two weeks period, a variety of weather regimes that are representative for Central Europe. We divide the forecasts into two categories based on the strength of the synoptic forcing, measured by the convective adjustment timescale. By doing so, we can investigate the regime dependence of predictability of summer convection, and by dividing the domain as before, the interconnection of synoptic weather regime and orography.

In order to objectively infer the predictive skill of the various ensemble experiments, we applied a hierarchy of measures ranging from visual inspection over NRMSE on different scales to the decorrelation and believable scale. The metrics chosen are suited for verification on the convective scale and provide meaningful and interpretable results. Also, the same metrics are used over the range of model configurations applied in this dissertation to further facilitate interpretation.

In the following sections, the NWP model, DA algorithm and details of the configurations mentioned above will be introduced in more detail. The metrics and verification methods are described, and their abilities are compared in an idealized testbed. Furthermore, the computation of the convective adjustment timescale is demonstrated.

2.2. Forecast Model: COSMO

The forecasts in this study are performed with COSMO, a non-hydrostatic limited-area NWP model (Baldauf et al., 2011). It is designed for operational and scientific applications on the convective scale and used operationally at DWD in varying configurations since April 2007. COSMO is based on the primitive thermo-hydrodynamical equations describing compressible flow in a moist atmosphere, which describe the external impact of gravity and Coriolis force, as well as the internal heat, mass and momentum transfer and phase changes of water:

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla p + \rho \mathbf{g} - 2\boldsymbol{\Omega} \times (\rho \mathbf{v}) - \nabla \cdot \underline{\mathbf{t}} \quad (2.1)$$

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{v} \quad (2.2)$$

$$\rho \frac{dq^x}{dt} = -\nabla \cdot \mathbf{J}^x + \mathbf{I}^x \quad (2.3)$$

$$\frac{de}{dt} = -p \nabla \cdot \mathbf{v} - \nabla \cdot (\mathbf{J}_e + \mathbf{R}) + \epsilon \quad (2.4)$$

The symbols represent time t , pressure p , density of the air mixture ρ , specific internal energy e , barycentric velocity \mathbf{v} , diffusion flux of internal energy (heat flux) \mathbf{J}_e , flux density of solar and thermal radiation \mathbf{R} , stress tensor due to viscosity $\underline{\mathbf{t}}$, kinetic energy dissipation due to viscosity ϵ , constant angular velocity of earth rotation $\boldsymbol{\Omega}$, apparent acceleration of gravity \mathbf{g} , and partial density of the mixture ρ^x , mass fraction q^x , sources and sinks \mathbf{I}^x and diffusion flux \mathbf{J}^x of constituent x .

The model equations are discretized on a rotated geographical coordinate system and a terrain-following Gal-Chen vertical coordinate and solved with a second-order Runge-Kutta scheme. A wide range of physical processes, unresolved by the grid, are accounted for by parameterization schemes. In all experiments of this study, the interactive multi-layer soil model (TERRA-ML) with seven soil layers and surface friction is used, as well as the single-moment bulk microphysics scheme accounting for cloud water, rainwater, cloud ice, snow, and graupel (Lin et al., 1983; Reinhardt and Seifert, 2006). We also use the operational two-stream radiation scheme (Ritter and Geleyn, 1992). Further details regarding discretization, parametrizations, and numerics can be found in the COSMO Core Documentation (Baldauf et al., 2018).

2.3. Data Assimilation Algorithm: KENDA

We used the KENDA system operational at DWD (Schraff et al., 2016), which is based on the local ensemble transform Kalman filter (LETKF) as described in detail by Hunt et al. (2007). In the following, the basic concepts of the Kalman filter (Kalman, 1960) and the Ensemble Kalman Filter (EnKF) (Evensen, 1994) are outlined to provide the reader with necessary background (compare Katzfuss et al., 2016).

Today, DA provides ICs, or analysis states, for all operational NWP systems by combining a forecast, often called first guess, with observations of the atmosphere. It does this in a sequential way, called cycling, where a forecast is started from an analysis providing a first guess for the next analysis and so on. Finding the optimal combination of first guess and observations in model space is equivalent to minimizing the following cost function:

$$J(x) = \underbrace{(\mathbf{x} - \mathbf{x}^f)^T \mathbf{P}^{f-1} (\mathbf{x} - \mathbf{x}^f)}_{\text{model part}} + \underbrace{(\mathbf{y}^o - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathbf{H}\mathbf{x})}_{\text{observation part}}, \quad (2.5)$$

where \mathbf{x}^f is the first guess, \mathbf{y}^o the observation vector and \mathbf{x} a model state. \mathbf{P}^f represents the error covariance matrix of the first guess, \mathbf{R} is the error covariance matrix of the observations and \mathbf{H} is the linear forward operator, which maps the model on the observation space. We assume a linear operator \mathbf{H} for clarity of nomenclature in all equations. The two parts of cost function 2.5, determining the analysis, can be nicely distinguished. The model part consists of the first guess (\mathbf{x}^f) containing information from all previously assimilated observations propagated in time by the model, which is imperfect due to the chaotic nature of the atmosphere (\mathbf{P}^f). The observation part includes all observations to assimilate in the current cycle (\mathbf{y}^o) and measurement uncertainties (\mathbf{R}), such as instrument or representation errors. Minimizing J yields the analysis state \mathbf{x}^a , which is equal to the right-hand side of the formulation of the Kalman filter (Lorenc, 1986):

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}[\mathbf{y}^o - \mathbf{H}\mathbf{x}_t^f] \quad (2.6)$$

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_t^f \quad (2.7)$$

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \quad (2.8)$$

$$\mathbf{x}_{t+1}^f = \mathcal{M}[\mathbf{x}_t^a]. \quad (2.9)$$

Here \mathbf{K} is the *Kalman Gain* and \mathcal{M} is the model to propagate the analysis in time. The first guess \mathbf{x}_t^f , valid at some time t , is combined with the information of the new observations to arrive at the analysis state \mathbf{x}_t^a with corresponding analysis-error covariance \mathbf{P}_t^a . The *Kalman Gain* is used to give appropriate weights to first guess and observations. Apart from minimizing Equation 2.5, those equations can also be derived by imposing that \mathbf{x}_t^a satisfies a linear combination of first guess and observation as in equation 2.6 for some \mathbf{K} , and \mathbf{K} is chosen in such a way that the analysis is unbiased and its error variance is minimized.

So far, we did not resolve how the first guess error covariance matrix \mathbf{P}^f is determined. The Ensemble Kalman Filter is able to take uncertainties into account by means of an ensemble that can provide a flow-dependent estimate of $\mathbf{P}^f = (N_{ens} - 1)^{-1} \mathbf{X}^f \mathbf{X}^{fT}$. $\mathbf{X}^{a,f}$ denotes the differences between (analysis/first guess) ensemble members $\mathbf{x}^{a,f}$ and (analysis/first guess) ensemble mean $\bar{\mathbf{x}}^{a,f}$ and N_{ens} is the ensemble size. This flow-dependence is a key capability of the LETKF and can lead to better performance in convective-scale applications compared to other algorithms, like 3DVar. Houtekamer et al. (2005) and Hamill and Whitaker (2011) confirm that forecasts initialized from EnKF analyses produce better probabilistic predictions compared to forecasts initialized by random stochastic perturbation. We distinguish two different approaches to create ensembles—*stochastic* and *deterministic* Kalman filters. In a *stochastic* Kalman filter, the observations are disturbed randomly in each ensemble member based on \mathbf{R} in such

a way that equation 2.7 is satisfied as $N_{ens} \rightarrow \infty$. Thereby for a finite ensemble, small and spurious correlations between the first guess and observations can be introduced, potentially resulting in a degraded analysis.

For *deterministic* Kalman filters, it is assumed that the optimal Kalman Gain is available and therefore the analysis error covariance is obtained from equation 2.7. Then the analysis is computed for the ensemble mean and in a subsequent step an ensemble of analysis perturbations \mathbf{X}^a is created in such a way that equation 2.7 is satisfied. This procedure is a key difference to other sequential Monte-Carlo algorithms, like particle filters, which rely on a reweighting step that can lead to filter collapse in high-dimensional problems (Snyder et al., 2008). This step is referred to as the ‘ensemble transform’ in the LETKF equations as they are applied in KENDA (Houtekamer and Zhang, 2016):

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{X}^f \tilde{\mathbf{P}}^a \mathbf{Y}^{fT} \mathbf{R}^{-1} (\bar{\mathbf{y}}^o - \mathbf{H} \bar{\mathbf{x}}^f) \quad (2.10)$$

$$\tilde{\mathbf{P}}^a = [(N_{ens} - 1)\mathbf{I} + \mathbf{Y}^{fT} \mathbf{R}^{-1} \mathbf{Y}^{f-1}] \quad (2.11)$$

$$\mathbf{X}^a = \mathbf{X}^f [(N_{ens} - 1)] \tilde{\mathbf{P}}^a]^{1/2}, \quad (2.12)$$

where \mathbf{Y}^f denotes the first guess perturbations in observation space. It can be shown that the choice of \mathbf{X}^a in equation 2.12 yields equation 2.7. $\tilde{\mathbf{P}}^a$ has dimension $N_{ens} \times N_{ens}$, making the computations very efficient. Additional efficiency is accomplished due to the fact that a time offset between the validity of an observation and the analysis can be corrected (Hunt et al., 2004) and the weights can be interpolated in space (Yang et al., 2009) so that the analysis can be computed on a coarser grid than the model. Mitchell and Houtekamer (2009) showed that a deterministic filter can obtain an analysis of similar quality with a smaller ensemble than a stochastic filter, thus further reducing computational demands.

In operational NWP, the ensembles are typically still too small to represent the atmospheric variability, and spurious correlations in \mathbf{P}^f are introduced due to sampling errors. Therefore, some form of localization has to be applied, meaning that correlations between distant points are neglected due to their unphysical origin. The LETKF performs local analyses on subspaces of the observation space instead of a single global analysis. The size of the subspace already excludes too distant observations. Additionally, we assign the remaining observations a weight based on their distance by localizing the observation error covariance matrix \mathbf{R} with a Gaspari-Cohn function (Gaspari and Cohn, 1999).

This split in local areas in observation space is exploited to create another significant advantage of the LETKF (Hunt et al., 2007). If the analysis is performed locally, the ensemble members only have to represent the atmospheric uncertainty in a subspace. As the linear combination of ensemble members can be different for each subspace, the resulting global analysis is not confined to the k -dimensional ensemble space, but explores a much more high-dimensional space—a fact of seminal importance, especially

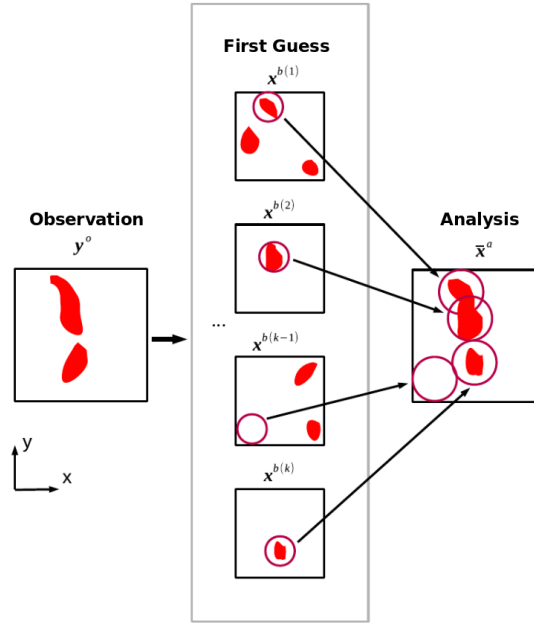


Figure 2.1: Conceptual visualization of an LETKF: horizontal cross-sections of a variable, for example precipitation, are shown for the observation and the first guess ensemble. The arrows indicate the local linear combinations that are used to build up a global analysis state. The figure is taken from Lange (2016).

on convective scales with their high spatio-temporal variability. The computational cost of the Matrix inversion in equation 2.8 is also reduced if independent, local analyses, which also facilitate parallelization, are computed and combined to one global analysis afterward.

In conclusion, the LETKF is a local, deterministic, square-root (see Eq. 2.12) Ensemble Kalman Filter, providing a linear combination of the first guess ensemble members, which is optimal in a least-squares sense in model space given the model uncertainty and the observation errors. In the KENDA system, all prognostic variables of the model are updated in the analysis (namely: zonal, meridional and vertical wind u, v, w , temperature T , pressure perturbation PP , mixing ratio of water vapor QV , cloud water QC , cloud ice QI , rainwater QR , snow QS , and graupel QG).

In Figure 2.1, the concepts introduced above are visualized in a simplified, schematic way. It shows horizontal cross-sections of, for example, precipitation in the observation, first guess ensemble and the resulting analysis. The LETKF analysis \bar{x}^a is the linear combination of first guess ensemble members \mathbf{x}^f , whereby the local weights are determined by how well the observations \mathbf{y}^o are matched. In this example, the observed squall line is put together from single convective cells in the first guess ensemble, resulting in an analysis that is much closer to the observations than any single first guess member. One can also immediately anticipate one of the issues of the LETKF. If the observation is outside of the ensemble span, meaning none of the ensemble members reproduces an observed convective cell, the LETKF is unable to improve the analysis in this respect.

2.4. Experimental Setup

We examine the impact of different factors on the predictability of convective precipitation, using different configurations of a convection-permitting NWP system. The experiments are all based on COSMO (Baldauf et al., 2011) coupled with KENDA (Schraff et al., 2016)), which is based on the LETKF (Hunt et al., 2007).

First, an idealized, perfect model, OSSE version of COSMO with assimilation of *synthetic* radar observations is used to investigate the impact of radar DA, and orography in a controlled perfect model environment that reduces the overall complexity of the system and provides a bias-free *truth*. Furthermore, we compare these impacts in a setting with only small-scale IC perturbations and one with added synoptic-scale perturbations to represent boundary condition (BC) errors of a LAM. The setup was successfully applied in several studies (Lange and Craig, 2014; Bachmann et al., 2019b,a).

Secondly, we investigate forecasts over Germany with the state-of-the-art COSMO-KENDA system (operational from March 2017 to May 2018¹) for a two weeks period in summer 2016 with recurring, heavy convective precipitation over Germany, in a realistic setting allowing us to generalize our findings. The weather situation in this period is comparable to the case in the idealized simulations, but exhibits more natural variability. As the experiments are computed in a realistic domain over Germany, complex orography is introduced. For this period, we compare experiments with LHN (Stephan et al., 2008) and direct radar data assimilation, as well as two different settings of the model mixing length.

Finally, COSMO-DE-EPS forecasts for three consecutive summer periods with the then operational NWP system are analyzed. Throughout three summers the weather situation exhibits more variability and not only convective, but also frontal precipitation. In this extensive data set, we can distinguish two weather regimes that are defined by their large-scale forcing of the convection. We can also compare the forecasts to radar-derived observations in a statistical sound sense. All model configurations and their specifications are summarized in Table 2.1 and will be introduced in more detail in the following.

2.4.1. Idealized Model Setup

For parts of the present study, COSMO-KENDA is modified to run in an idealized OSSE setup (COSMO version 5.3). By this, the true state of the atmosphere is known, which allows disentangling different uncertainties. In contrast to the now operational setting,

¹since then the domain size and horizontal resolution are increased

observations of radar reflectivity and wind velocity are assimilated directly. We do not use LHN, but instead, we simulate *synthetic* radar reflectivity and velocity observations from a Nature Run and assimilate them directly in KENDA. In an exploratory study, Bick et al. (2016) tested the direct assimilation of radar observations over Germany in COSMO-KENDA with potentially beneficial results.

Nature Run and Background Ensemble

The Nature Run comprises a deterministic COSMO forecast and represents the *true state* of the atmosphere. At a convection-permitting horizontal resolution of 2 km, the domain encompasses $256 \times 256 \times 50$ grid points with periodic lateral boundaries. Due to the periodic boundaries, the domain evolves as a whole and is not steered by an inflow of another model. In the vertical, a terrain-following Gal-Chen coordinate system with a layer thickness ranging from ~ 100 m near the surface to ~ 800 m at the top of the domain is used. This results in a total domain size of $512 \times 512 \times 22$ km, which was chosen to be a multiple of two to make verification easier, especially in spectral space and coarse-graining. The Coriolis force is not included in the calculations, and neither deep nor shallow convection is parameterized, but explicitly resolved to increase the realism of the convective representation.

The Nature Runs (called IDEAL-Nat hereafter), one with orography (IDEAL-Nat *oro*) and one without (IDEAL-Nat *flat*), are started from idealized, horizontally homogeneous ICs based on a sounding from Payerne in Switzerland (CH, Radiosonde 06610), valid at 1200 UTC on 30 July 2007 (Bischof, 2011; Lange and Craig, 2014; Bachmann et al., 2019b). It features high Convective Available Potential Energy (CAPE) of $\sim 2200 \text{ J kg}^{-1}$ and vertical wind shear, which allows for strong and long-lived convection. The mean wind of about 5 m s^{-1} at 1500 m above ground is from the South-West (see Fig. 3.1). Therefore, the pure advection of an object, e.g., a thunderstorm, through the entire model domain would take about 20 h, which is much longer than our forecast range or the lifetime of convection in the experiments thus eliminating the possibility of multiple interactions between orography and thunderstorm. In combination with the simple orography, this facilitates a clear attribution to orographic effects.

We perturbed those homogeneous ICs in two ways: First, by adding white noise with an amplitude of 0.02 K on the temperature field and 0.02 m s^{-1} on the vertical velocity field within the lowest 100 hPa of the atmosphere to break the initial symmetry; and second, with height-correlated errors (0.25 K on the temperature, 0.25 m s^{-1} on the horizontal wind and 2% on the relative humidity), which represent synoptic-scale uncertainties, on both ICs and lateral boundary conditions (LBCs), in reality emerging from boundary conditions of a LAM (referred to as boundary condition perturbations in the following).

The observational sounding and the extent of the synoptic-scale perturbations is displayed in Fig. 2.2a. The perturbations on the wind are rescaled to ensure that the total wind speed remains constant and only the wind direction is modified in order to reduce complexity. For physical reasons, perturbations on the relative humidity are ensured not to cause supersaturation. Figure 2.2b provides a scatter plot of CAPE and

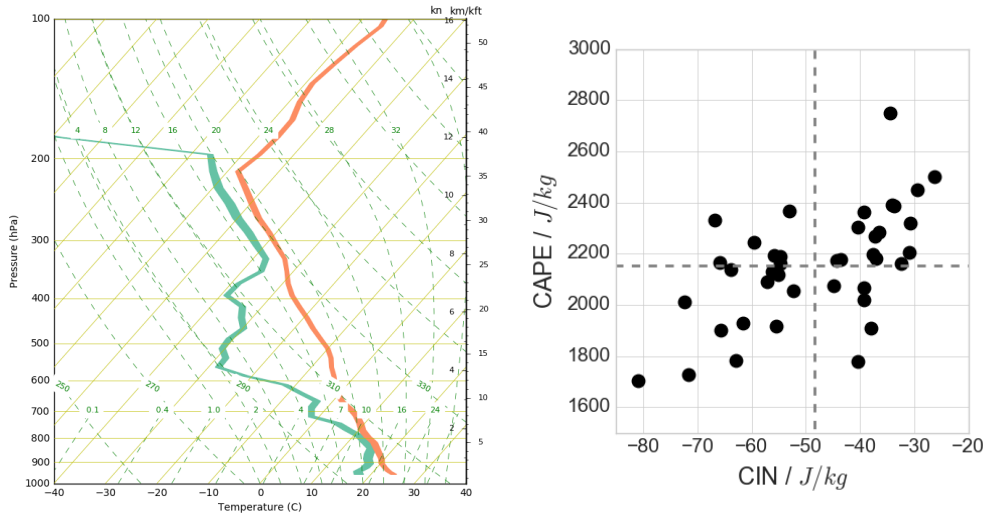


Figure 2.2: (a) Initial sounding with the added boundary condition errors. The envelope of the ensemble temperature is shown in orange and the dew point temperature in green. (b) Variability of CAPE and CIN in the initial sounding of the IDEAL IBC experiments.

Convective Inhibition (CIN) of the ensemble members and visualizes the differences due to those perturbations. CAPE ranges from 1700 to 2800 J kg^{-1} and CIN from -80 to -25 J kg^{-1} . These variations lead to more realistic variability in the resulting convection and its timing (cf. Cintineo and Stensrud, 2013), visible in Fig. 3.6 and Fig. 3.7 and further discussed in Section 3.1.2.

A background ensemble with 40 members (as recommended in Harnisch and Keil, 2015), called IDEAL, without any data assimilation serves as the reference for the experiments with radar DA. It is started at 0400 UTC in the morning and integrated for 10 h (see Fig. 3.2) with either only IC perturbations (referred to as IC) or initial and boundary condition perturbations (referred to as IBC). All experiments and their specifications are also summarized in Table 2.1.

The orography is represented by a single, point-symmetric Gaussian mountain with 1000 m height and a half-width of 10 km located at $x, y = (64 \text{ km}, 64 \text{ km})$. All experiments are performed in two sets, one without (*flat*) and one with orography (*oro*) to allow for comparison. The height, as well as the half width, is chosen to obtain a significant elevation of about 150 m at a radius of 20 km, a scale at which Robinson et al. (2008) and Cronin et al. (2015) found the strongest deep-convective response. As Picard and Mass (2017) found almost no sensitivity to the shape of the orography, we limit our investigation to a Gaussian mountain shape.

Radar DA and Synthetic Observations

For the idealized experiments, COSMO-KENDA is modified to run in an idealized OSSE setup. This way, the true state of the atmosphere is known, which allows disentangling different uncertainties. We do not use LHN, but instead, we simulate *synthetic* radar

reflectivity and velocity observations from the Nature Run and assimilate them directly in KENDA, which is modified to produce a fully periodic analysis.

Only conventional observations and aircraft data in the vicinity of airports (Mode-S, Lange and Janjić, 2016) are currently assimilated directly at DWD, whereas radar observations are incorporated using LHN. In an exploratory study, Bick et al. (2016) tested the additional direct 3-dimensional assimilation of radar data in COSMO-KENDA and showed potential benefits. Snook et al. (2011, 2012) report beneficial impact on the analysis and forecast of a configuration similar to the one in this study in the sense that only radar data and no conventional observations are assimilated.

The role and impact of IC uncertainties are examined by assimilating radar observations simulated from the Nature Run, following Done et al. (2004) and Lange and Craig (2014). The reflectivity a radar would measure is calculated with a comparatively simple operator based on the Z-R relation, using the mixing ratios of rain, snow, and graupel on the model grid at 2 km horizontal resolution with a vertical resolution of 1 km, reaching from 500 m to 14.5 km above ground. Gaussian noise with a standard deviation of 5 dBZ is added to simulate measurement errors. The zonal wind component, with added noise (standard deviation of 1 m s^{-1}), serves as a proxy for observations of radar radial velocity. Radar velocity observations are limited to regions where reflectivity is larger than 5 dBZ. Although the observation errors are relatively small compared to measurement errors of radar systems, they are in line with previous studies. Snook et al. (2015) use error standard deviations with 3 dBZ for radar reflectivity and 2 m s^{-1} for radial velocity (2 dBZ and 1 m s^{-1} in Snook et al., 2011).

Radar reflectivity observations are thresholded at 5 dBZ in such a way that all data below that threshold is considered as ‘zero precipitation’ or ‘no reflectivity’ (cf. Aksoy et al. (2009)). The assimilation of ‘zero precipitation’ observations suppresses so-called false positives, precipitation in the forecast that is not observed. The observations are coarsened to super-observations with a resolution of 8 km, close to the effective resolution of COSMO ($5\Delta x \sim 10 \text{ km}$, Bierdel et al., 2016). In general, there are two options for super-observations: Thinning and averaging. In the present study, super-observations are generated by spatially averaging over tiles of 4×4 grid points, as less information is discarded in this case.

KENDA typically uses adaptive covariance inflation; in contrast it is set to a constant value of 1.2 in this study to allow for a fair comparison of the experiments. The horizontal localization is constant as well (32 km), while the vertical localization increases with height in the range from 0.075 to 0.5 in terms of the logarithm of pressure as in the experiments of Schraff et al. (2016). The horizontal localization is chosen to be in the range of the assimilated feature, in our case convection, and comparable to previous studies (e.g., 28 km in Cintineo et al., 2016). The cycling interval is 15 min, which probably will be shorter than operational applications for a couple of years, but takes into account the rapid convective development due to latent heat release. It is also used in Lange and Craig (2014) and Bick et al. (2016).

The configuration of the DA cycling in the idealized experiments is schematically shown in Figure 2.3. After a 3 h spin-up period, the synthetic observations, provided by the

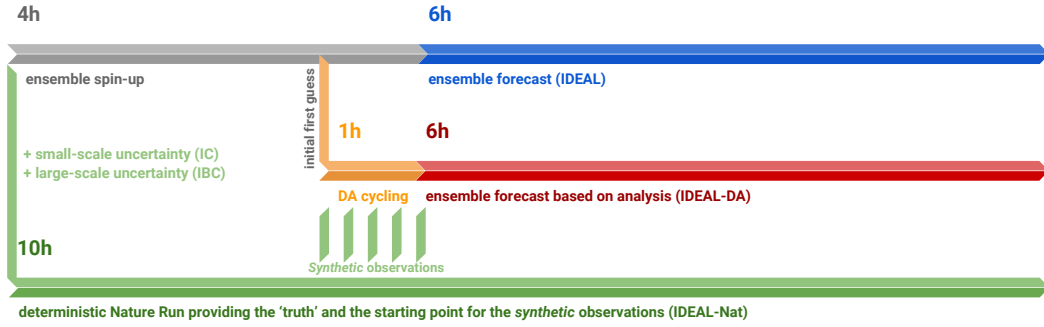


Figure 2.3: Flow diagram of experimental setup in idealized experiments: the deterministic Nature Run provides the *synthetic* observations (green), which are combined with the first guess ensemble (grey) in the DA cycle (orange). IDEAL-DA (red) is initialized by the last analysis and IDEAL (blue) provides a background ensemble without DA for comparison.

Nature Run, are assimilated every 15 min for a period of 1 h around the onset of convective precipitation between 0700 UTC and 0800 UTC. We then perform several ensemble forecasts (40 members) with 6 h forecast lead time issued from the last analyses valid at 0800 UTC.

The forecasts shown in the flow diagram (see Fig. 2.3) have to be computed independently for *flat* and *oro* and for IC and IBC, which results in four sets of simulations. The sensitivity experiments introduced in section 2.4.1 only differ in the configuration of KENDA and therefore IDEAL and IDEAL-Nat can be reused, while the cycling and ensemble forecasts are recomputed (see Table 2.1 for a comprehensive summary of the abbreviations).

Sensitivity Experiments

Four sets of ensemble experiments are conducted to investigate various sensitivities. First, the hourly DA window is shifted by 4 h (to 1100 – 1200 UTC) to examine the impact of radar DA at a later time within the life cycle of convection (called DAtime).

Secondly, the sensitivity regarding the magnitude of the error assigned in the DA is tested. Observation errors of radar reflectivity and velocity are tripled (15 dBZ and 3 m s^{-1}) in IDEAL-DAobserr3, resulting in significantly lower weights of the observations in the DA system, which might be closer to a real system that faces issues of observation representativity and correlated observation errors that are often accounted for by inflating the assigned observation error (e.g. Bormann et al., 2003; Bauer et al., 2010). In another experiment, IDEAL-DAobserr05, the observation errors of radar reflectivity and velocity are taken by half (2.5 dBZ and 0.5 m s^{-1}), allowing us to estimate the impact of improved measurement networks or DA algorithms. The assimilated observations and their errors (white noise; standard deviation 5 dBZ) remain unchanged compared to the reference IDEAL-DA.

Thirdly, the sensitivity to the type of the assimilated observations is investigated in the experiments DAwind and DAreflectivity. In the DAwind experiment exclusively

radar velocity observations and in DAreflectivity only the reflectivity observations are assimilated. Whereas radar reflectivity observations necessitate a complicated nonlinear forward operator to depict them in model space, the assimilation of radar velocity is more accessible to implement as the horizontal wind is a direct model variable. The experiment DAwind, therefore, represents the current operational procedure at several NWP centers (see Gustafsson et al. (2017) for further details on operational procedures). DAreflectivity is performed for the comparison of the relative potential impact of reflectivity and velocity observations.

Table 2.1: Summary of abbreviations used to describe the experiments.

Abbreviation	Description
<i>flat</i>	Flat experiment without orography
<i>oro</i>	Orography experiment with Gaussian-shaped mountain of 1000 m height
<i>well-organized</i>	experiment based on ensemble member with organized convection
<i>popcorn-like</i>	experiment based on ensemble member with popcorn-like convection
IC	only initial condition errors
IBC	initial and additional boundary condition errors
IDEAL-Nat	deterministic Nature Run representing the truth, i.e. providing observations
IDEAL	idealized COSMO-KENDA ensemble forecast without any data assimilation
IDEAL-DA	idealized COSMO-KENDA ensemble forecast started from analysis using velocity and reflectivity observations in DA
IDEAL-DAtime	Ensemble forecast started from assimilation window, time-shifted by 4 h to 1100 UTC – 1200 UTC
IDEAL-DAobserr05	Ensemble with halved assigned observation errors for reflectivity and velocity (to 2.5 dBZ and 0.5 m s^{-1})
IDEAL-DAobserr3	Ensemble with tripled assigned observation errors for reflectivity and velocity (to 15 dBZ and 3 m s^{-1})
IDEAL-DAwind	Ensemble forecast from analysis using only velocity observations in DA
IDEAL-DAreflectivity	Ensemble forecast from analysis using only reflectivity observations in DA
DE-DA-LHN	COSMO-KENDA ensemble with latent heat nudging valid for a high impact weather period of 14 days
DE-DA-3DRad	COSMO-KENDA ensemble with direct assimilation of 3D reflectivities and radial velocity and warm bubbles triggered by analysis of missing reflectivities in the model valid for a high impact weather period of 14 days
DE-EPS	COSMO-DE-EPS ensemble over 3 summer periods between 2014-2016

2.4.2. Operational Model Setup

We also analyzed ensemble forecasts of different configurations of an operational or pre-operational COSMO setup, step by step reintroducing complexity. Those are combined with various techniques to generate an ensemble, which are KENDA, LHN and a down-scaling approach. We describe those in the following.

DE-DA setup: High Impact Weather period

In order to simulate convection in a more realistic environment, COSMO (in version 5.4h) was used to produce forecasts of a convectively very active period that is similar to the situation in the idealized experiments. It provides forecasts in a domain centered over Germany (10°E and 50°N , see Fig. 2.5). The domain encompasses 461×421 grid points, with a horizontal grid spacing of 0.025° or roughly 2.8 km. The same vertical coordinates and the same parametrizations as in the idealized setup are applied to minimize differences in the model configuration.

In the present study, 6 h forecasts of a 20 member ensemble are initialized every hour between 1000 UTC and 1800 UTC every day between 27 May and 10 June 2016, resulting in the 126 ensemble and 2520 individual forecasts.

Technically, the model setup is very similar to the idealized setup described in Section 2.4.1, apart from the model domain and increased natural variability in the weather situations investigated. However, there are some differences in the DE-DA setup compared to IDEAL-DA.

The first difference is the mixing length, a model parameter connected to convective triggering. It describes the characteristic length scale for vertical mixing in the boundary layer scheme (Baldauf et al., 2011) and has been set to 150 m in the operational setup since 2008. However, it has been discussed that a value of 500 m is more realistic and the assimilation of 3D radar data might benefit from changing the mixing length from 150 to 500 m. Those values are also applied and discussed in Harnisch and Keil (2015). A smaller mixing length allows for the boundary layer to heat up more, which results in prior and deeper convection but also causes a temperature and surface pressure bias (Hanley et al., 2015; Necker et al., 2018; Hirt et al., 2019, submitted). In the present study, we will discuss the findings for both settings.

The second significant difference concerns the DA cycling. The DA system is cycled with 40 ensemble members, using 1 h assimilation windows from 26 May 2016, 0000 UTC until 10 June 2016 (blue in Fig. 2.4). The LBCs in the assimilation cycle for each first guess integration are provided by the most recent ensemble forecasts of the Icosahedral Nonhydrostatic Model over European Union (ICON-EU) (green in Fig. 2.4), which are issued every 3 h. The 6 h forecasts, initialized hourly between 1000 UTC and 1800 UTC,

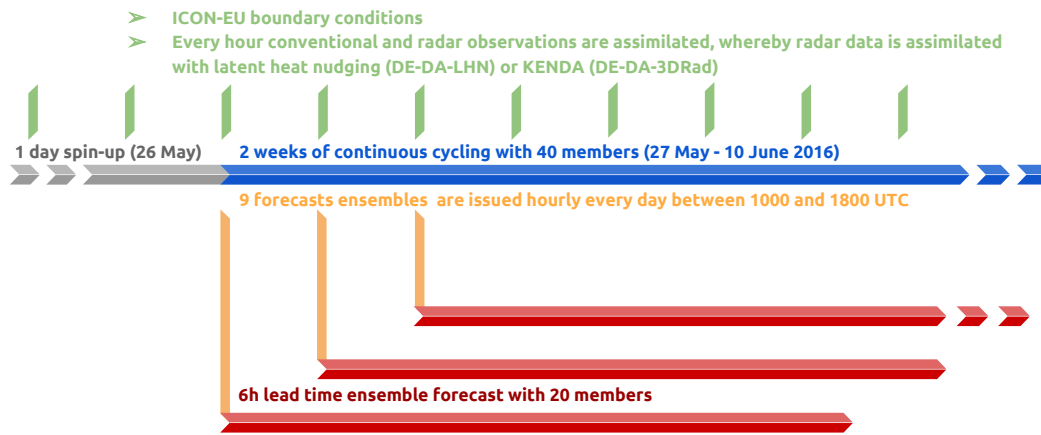


Figure 2.4: Flow diagram of the cycling in the HIW period: a 40 member background cycling (blue) based on ICON-EU boundary conditions and either radar DA (DE-DA-3DRad) or LHN (DE-DA-LHN) (green) provides the analyses (orange) to start the forecast ensembles with 20 members (red). The initial spin-up period is indicated in grey.

are initialized from the corresponding analyses starting from 27 May 2016 (orange in Fig. 2.4), which ensures one day of spin-up for the whole system (grey in Fig. 2.4).

This is necessary as the first guess ensemble transports information from previous observations into the DA cycling. Therefore, the forecast skills of early analyses that are based on fewer observations are less skillful. The complete setup is schematically visualized in Fig. 2.4. The 40 background members (blue in Fig. 2.4) ensure a high quality of the analysis, as they are used to sample the flow-dependent error covariance of the atmosphere with the ensemble. In contrast, the forecasts (red in Fig. 2.4) are not used as first guesses at later times and can be computed with 20 members to reduce computational cost.

The radar (reflectivity) observations are assimilated in two different ways, which will be explained in the following, while the assimilation of conventional data is identical in both simulations. The same combination of different inflation methods is used to create the required spread in the ensemble. We apply adaptive, multiplicative covariance inflation and the relaxation to prior perturbations (RTPP, Schraff et al., 2016). Additionally, model errors are added to the analysis ensemble members, using the climatological error covariance matrix \mathbf{B} of the global 3DVar system as a proxy. The DE-DA-LHN setup was used for operational NWP at DWD from March 2017 to May 2018 before the model domain was extended. Apart from the domain, KENDA is still running with the same settings. The DE-DA-3DRad setup includes several features that are planned to be applied in future DWD systems and that are developed within the Seamless INtegrated FOrcastiNg sYstem (SINFONY) project at DWD.

Latent Heat Nudging (DE-DA-LHN)

The first approach uses LHN in the analysis cycle of every ensemble member, but not

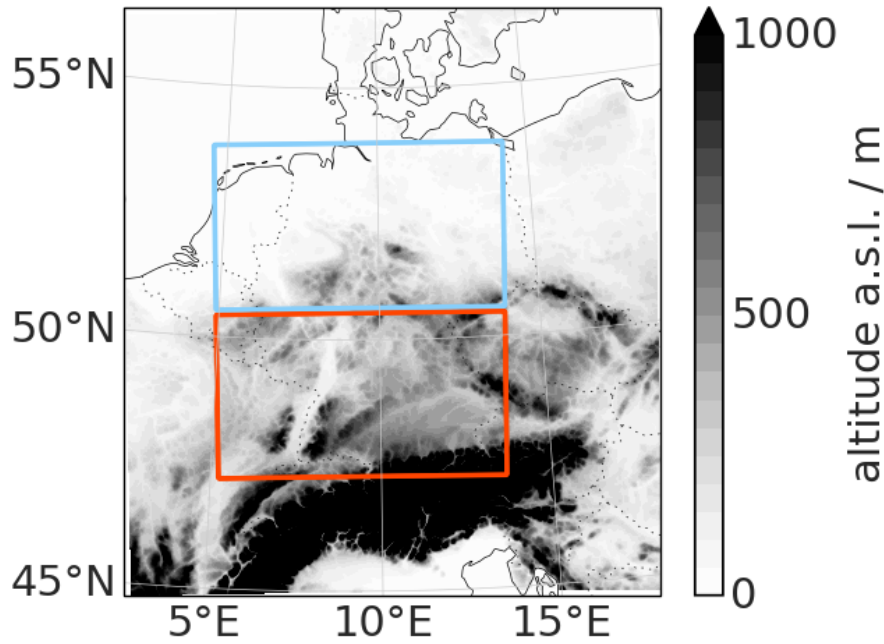


Figure 2.5: Map of the investigated COSMO-KENDA and the DE-EPS domain: the altitude above sea level is shown in grey-scale and the blue (red) box indicates the sub-domain *North* (*South*); political borders are displayed in dotted lines for orientation.

the free forecasts. LHN is applied during the model integration at every time step, using interpolated radar data from the precipitation scan (available every 5 min), while we discard the radar wind observations. This setup is called DE-DA-LHN hereafter. LHN is one approach to initialize convection, based on surface precipitation rates derived from radar reflectivity data (Stephan et al., 2008). It was introduced in 2007 and is still used operationally. Conceptually, LHN adds a heating term to the prognostic equations of the model if the model produces precipitation that is not matching the observations at a specific location. If precipitation is underestimated, the term will be positive, thereby representing latent heat release in a convective cell.

3D Radar DA (DE-DA-3DRad)

In this setup, model equivalents of synthetic 3D radar data are obtained, using the Efficient Modular VOlume scanning RADAr Operator (EMVORADO, Zeng et al., 2014, 2016). The assimilation of reflectivity and the radial wind is performed following Bick et al. (2016). Additionally, warm bubbles can be triggered independently for each ensemble member every 15 min during the model integration to account for observed convective cells missing in the model representation at the corresponding time (see Zeng et al., 2019).

COSMO-DE-EPS: 3 Years of Operational Forecasts

COSMO-DE-EPS provided forecasts over Germany in the same domain as DE-DA (see Fig. 2.5), using a combination of physics parameter perturbations and boundary conditions, down-scaled from four different global EPS, to generate an ensemble (Gebhardt et al., 2011; Peralta et al., 2012; Ben Bouallègue and Theis, 2014; Kühnlein et al., 2014). The physics parameters—for example, the already mentioned mixing length or the entrainment rate for shallow convection—are perturbed in a non-stochastic and uniform way, so the perturbation is constant over the whole forecast, providing a crude estimate of model error. The global models are the Global Model Erde (GME) of DWD, the Integrated Forecasting System (IFS) of ECMWF, the Unified Model of UK MetOffice and the Global Forecasting System (GFS) of National Centers for Environmental Prediction (NCEP). The ICs are derived from COSMO-EU (horizontal resolution 7 km) and the COSMO-DE deterministic run, which includes radar observations assimilated via LHN, but no direct assimilation of reflectivity or radial wind observations.

It features a more straightforward approach to generate an ensemble, which was nevertheless successfully used operationally for several years, starting in May 2012. The ensemble forecasts (20 members) are initialized at 0000 UTC and 1200 UTC for a lead time of 27 h in the summer periods of 2014–2016 (May–September). This long period includes various weather situations. The horizontal grid spacing of 2.8 km again allows for the explicit representation of deep convection, while shallow convection is parameterized in this configuration. Other than that, the same parametrizations as in the different setups are applied, easing comparability.

Observations for Verification

We compare the forecasts to the so-called EY product of DWD, which provides radar-derived accumulated precipitation over Central Europe with a horizontal resolution of 1 × 1 km and a time resolution of 5 min. The high resolution is seminal due to the rapid spatial and temporal development of convection. The EY product was coarse-grained to the COSMO grid, and the precipitation was accumulated hourly to facilitate comparisons to the forecasts. The EY product covers the entire investigated domain of this study.

For further investigations, the domain is split in a Northern and a Southern part, highlighted by the blue and red boxes in Fig. 2.5. As indicated by the grey-scale, less orography is present in the Northern half, while large parts of the Alps, Vosges, and the Mittelgebirge are within the Southern half, making the division in an orographically influenced Southern part and a comparatively plain Northern part possible.

2.5. Metrics and Scores

In order to objectively infer the predictive skill of ensemble experiments, we applied a hierarchy of measures ranging from visual inspection over the Normalized Root Mean Square Error (NRMSE) on different scales to the believable and decorrelation scale.

To assess the skill and predictability of the forecasts, we use a range from a conventional measure, as the NRMSE, to more sophisticated scores, as the FSS and the decorrelation scale. The chosen metrics are complementary, as the NRMSE penalizes intensity errors, while the believable scale, based on the FSS, as a neighborhood method, assesses the similarity of the forecasts to the truth on a certain scale, or in other words, the model predictability of the atmospheric state.

In contrast, the decorrelation scale is calculated in spectral space, includes no comparison to the truth and represents a scale-separation method. It can provide insights into the predictability of the model state, a proxy for intrinsic predictability. Scale-separation methods not only allow for displacements as neighborhood approaches, but also evaluate the skill at different scales independently (Casati et al., 2004). Also, all metrics are proper as they do not allow for a systematic hedging (Gneiting and Raftery, 2007; Gneiting and Katzfuss, 2014).

2.5.1. Convective Adjustment Timescale

In order to separate the forecasts of the investigated three years period, we apply the convective adjustment timescale τ_c . It is an indicator for the prevailing weather regime based on how fast conditional instability (measured by CAPE) is removed from the atmosphere by moist convection (measured by precipitation). A convective timescale that is much smaller than the timescale, on which the synoptic-scales develop, indicates an equilibrium state. Deep convection is therefore predominantly controlled by the synoptic-scale forcing. In contrast, if τ_c is large, no such synoptic forcing is present, and the convective instability depends on a local triggering mechanism to be released. The metric has been successfully applied as an indicator of the prevailing weather regime in previous studies (e.g. Done et al., 2006; Zimmer et al., 2011; Surcel et al., 2017).

The domain average of the convective timescale τ_c was computed for the hourly model output in the following way (Done et al., 2006):

$$\tau_c = \left(\frac{c_p \rho_0 T_0}{2L_v g} \right) \frac{\text{CAPE}}{P}. \quad (2.13)$$

All quantities within the bracket are constant (the specific heat of air at constant pressure c_p , the reference values of density ρ_0 and temperature T_0 , the latent heat of vaporization L_v and the gravity acceleration g) and only the CAPE (Joules/kg) and the precipitation rate ($P / \text{kg}(\text{sm}^2)^{-1}$) vary. In the present study, τ_c was only computed if a threshold of 100 grid boxes precipitated to exclude extreme outliers, and the forecasts are sorted into categories by applying a threshold of 6 h to the maximum daily value to distinguish between local and synoptic forcing situations. All results are robust to reasonable variations (3 h-12 h) of this choice.

2.5.2. Normalized Root Mean Square Error

The Normalized Root Mean Square Error (NRMSE) represents a widely used, continuous measure. The normalization becomes necessary as the RMSE is strongly correlated with the total amount of precipitation in the domain and therefore mainly depicts the diurnal cycle. The NRMSE is defined as follows:

$$\text{NRMSE}_{(n)} = \sqrt{\frac{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (O_{(n)i,j} - F_{(n)i,j})^2}{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (O_{(n)i,j} + F_{(n)i,j})^2}}, \quad (2.14)$$

where $O_{(n)}$ and $F_{(n)}$ represent the 2-dimensional observation and forecast fields and i, j the respective indexes. (n) indicates the side length of the square neighborhood, which is compared. Finally, NRMSE is averaged over all ensemble members. The NRMSE ranges from 0 to 1, where 0 would be a perfect forecast and values are typically smaller for shorter lead times. It is critical to note that the NRMSE is prone to so-called double penalties produced by phase errors in the forecasts. NRMSE has already been applied in previous studies (e.g. Surcel et al., 2014). In order to evaluate the NRMSE on different scales, a coarse-graining by $n \times n$ grid points is performed.

2.5.3. Decorrelation Scale

Surcel et al. (2015) defined a method to determine the scale-dependent predictability limit of an ensemble forecast. It is based on the variance of the sum of a variable Ψ_k :

$$\text{Var}\left(\sum_{k=1}^N \Psi_k\right) = \sum_{k=1}^N \text{Var}(\Psi_k) + \sum_{k \neq l} \text{Cov}(\Psi_k, \Psi_l) \quad (2.15)$$

If all variables Ψ_k and Ψ_l are fully decorrelated, then the covariance $\text{Cov}(\Psi_k, \Psi_l) = 0$ and the following ratio holds:

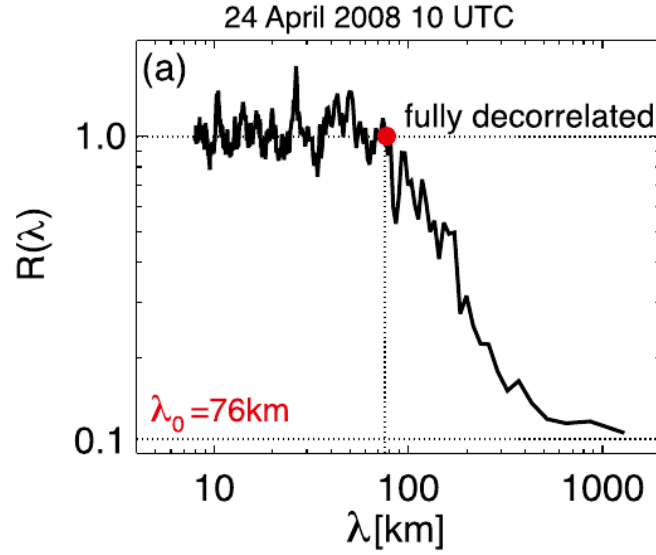


Figure 2.6: Illustration of the ratio $R(\lambda)$ defined in equation 2.16. The scale at which the ratio drops is highlighted by a red dot and marks the scale at which the ensemble becomes fully decorrelated, defined as the decorrelation scale. The figure is taken from Surcel et al. (2015).

$$\frac{\sum_{k=1}^N \text{Var}(\Psi_k)}{\text{Var}\left(\sum_{k=1}^N \Psi_k\right)} = 1, \quad (2.16)$$

where k represents the ensemble member and N the total ensemble size. Following Denis et al. (2002), the variance of a 2-dimensional field can be calculated by the Fast Fourier Transformation (FFT), which introduces a scale dependence into Equation 2.16. Therefore, a ratio $R(\lambda)$ can be computed for all wavelengths λ .

Figure 2.6 shows the ratio $R(\lambda)$ in a typical case. $R(\lambda)$ depicts values around one up to a certain scale and then drops rapidly towards $1/N$. The scale at which the ratio drops is highlighted by a red dot and marks the scale at which the ensemble becomes fully decorrelated. It is defined as the decorrelation scale. In practice, a threshold of 0.95 instead of 1 is used to account for noise. We assume that all smaller scales are unpredictable. A visualization of this method is a high-pass filter with a cut-off at the decorrelation scale, which yields a smoothed field containing only information within the ensemble and no noise.

In the present study, we used the FFT in all idealized experiments and the Discrete Cosine Transformation (DCT) for forecasts in the realistic domain since it provides superior aliasing in a limited domain without periodic boundaries. We compared the performance of FFT and DCT on the idealized domain thoroughly and found no significant differences in our results.

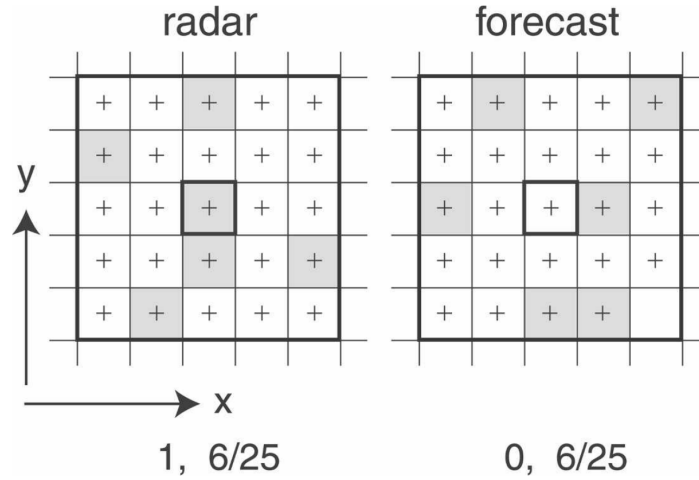


Figure 2.7: Illustration of the verification with a neighborhood approach. Shown is a hypothetical radar observation and forecast of precipitation, where precipitation is indicated by grey shading. The verification domain is shown by black boxes. The figure is taken from Roberts and Lean (2008).

2.5.4. FSS and Believable Scale

One way to overcome the double-penalty issue is so-called fuzzy verification or neighborhood methods (Gilleland et al., 2009). The FSS, introduced by Roberts and Lean (2008), is able to assess the spatial skill beyond a grid point level for deterministic and ensemble forecasts. As models are unable to represent flows below an effective resolution of approximately $5 \Delta x$ properly and a forecast of a slightly displaced thunderstorm might still be considered skillful by a forecaster, the FSS provides useful information. It is a commonly applied neighborhood method (Mittermaier and Roberts, 2010; Mittermaier et al., 2013; Schwartz and Sobash, 2017), whose characteristics are addressed in several publications (Skok, 2015, 2016; Skok and Roberts, 2018; Mittermaier, 2019).

Figure 2.7 illustrates the concept of a neighborhood approach. Verification of the forecast with radar observations on a grid scale level (central box) would yield a completely wrong forecast. If, however, a neighborhood (larger boxes) is considered, the forecast is perfect. By examining a neighborhood instead of grid points in verification, one relaxes the requirements in the forecast. It is deemed to be skillful, as long as it predicts the right area fraction of precipitation above a threshold on a particular scale, the neighborhood.

The FSS attains values between 0 (no skill) and 1 (perfect skill), where a score of $0.5 + f_0$ is defined as the limit to a skillful forecast and the respective neighborhood as skillful spatial scale or believable scale (Dixon et al., 2009; Mittermaier and Roberts, 2010), revealing the limit to scales of the forecast that contain no more skillful information but only random fluctuations. The fractional coverage f_0 , or the wet-area ratio, refers to the fraction of grid points above the threshold in the observation, as cases with larger

f_0 tend to provide greater skill because of a higher probability of overlap by chance. f_0 for the experiments in the present study is shown in Fig. 3.2. The believable scale is a direct way to identify skillful or predictable scales of the forecast. Therefore, we assume scales above the believable scale to be the predictable ones in the given setup.

The FSS is calculated by first creating binary fields with a threshold to observation and forecast. Afterwards, the fraction of grid points above the threshold within a square of length n surrounding every grid point of those fields is calculated by the nearest-neighbor method and compared to the Mean Square Error (MSE):

$$\text{MSE}_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left(\tilde{O}_{(n)i,j} - \tilde{F}_{(n)i,j} \right)^2, \quad (2.17)$$

where N_x and N_y are the number of horizontal grid points, $\tilde{O}_{(n)i,j}$ and $\tilde{F}_{(n)i,j}$ are the binary fields mentioned before and i, j represent the horizontal indexes. The score is calculated by scaling the result as follows:

$$\text{FSS}_{(n)} = 1 - \frac{\text{MSE}_{(n)}}{\text{MSE}_{(n),ref}}, \quad (2.18)$$

where $\text{MSE}_{n,ref}$ is the largest possible MSE for the forecasted and observed fractions. The believable scale is calculated by searching for the neighborhood size closest to the desired FSS score of $0.5 + f_0$ for every ensemble member and averaging the results. The believable scale is an extension to the FSS (Roberts and Lean, 2008). It is useful to compare forecasts with different model biases or unequal fractional coverage of precipitation, as it corrects those influences. It delivers a skillful scale above which the forecast resembles the truth more than a random realization of the precipitation coverage. As Faggian et al. (2014) laid out, those computations can be performed in a computationally very efficient way, which allows for a resolution of 2 km in the believable scale.

The treatment of lateral boundaries poses a challenge to all neighborhood methods, as window sizes must be constrained to a fraction of the maximum extent of the domain. In the idealized experiments in this study, we calculate the believable scale on the periodic domain and no further assumptions are required. In the real-world experiments, we pad the fields with zeros, as proposed in Roberts and Lean (2008).

In order to avoid confusion, we want to address the difference between the *believable* and the *skillful scale*, and the use of the term in this dissertation. The term *skillful scale* is used in studies that compare a deterministic or ensemble forecast to observations of the atmosphere, e.g. Roberts and Lean (2008) and Mittermaier et al. (2013). The *skillful scale* is a score in those studies. If the forecast is compared to another forecast or an idealized nature run without systematic model errors, the *skillful scale* generalizes to the *believable scale* (c.f. Dey et al., 2014, 2016). In those cases, it describes a level of spatial agreement, but is not a score. Strictly speaking, we show the *believable scale*

in Sec. 3.1–3.3 and the *skillful scale* in Sec. 3.4, but as the difference is strictly in the nomenclature, we use the term *believable scale* throughout this dissertation for clarity.

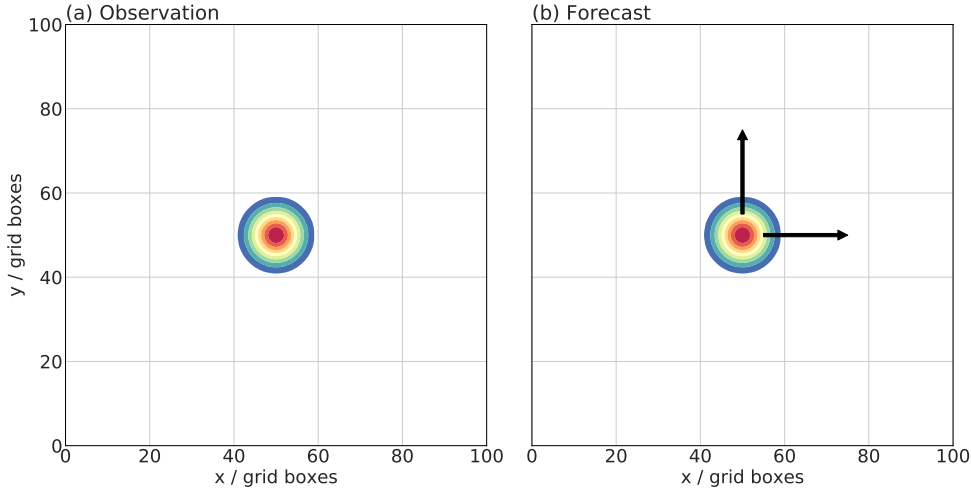


Figure 2.8: Illustration of the verification testbed used in this section: (a) shows the observation, a Gaussian representing convection, and (b) shows the respective forecast, a Gaussian that is shifted as indicated by the arrows.

2.5.5. Comparison of the Predictability Metrics

We want to give a brief overview, how the scores introduced above, compare to each other in a very idealized testbed to ease interpretation in the remainder of this thesis. We used a 100×100 grid point doubly-periodic domain to calculate the scores for well-defined forecasts of single Gaussian precipitation cells. We want to mention that similar, but more thorough investigations for the FSS have been published, which also give indications towards the behavior of the believable scale (Skok, 2015, 2016; Skok and Roberts, 2018; Mittermaier, 2019). However, those studies do not compare to other metrics.

In contrast to the NRMSE, the decorrelation and believable scale are scale-separation methods, which both allow for displacements in the forecasts and scale-dependent skill evaluation (Casati et al., 2004). All applied scores are proper as they do not allow for systematic hedging (Gneiting and Raftery, 2007). Besides, they complement each other as the decorrelation scale is solely based on the ensemble forecast, while the believable scale is a comparison to the observations. In other words, the decorrelation scale is a measure for the spatial spread or the predictability of the model state. NRMSE and believable scale are scores and assess the model predictability of the atmospheric state,

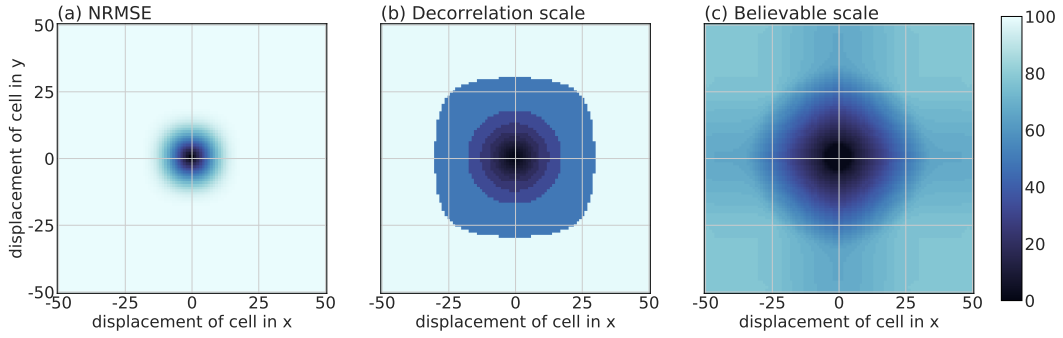


Figure 2.9: Verification of precipitation forecasts and the impact of forecast displacements with different metrics: (a) NRMSE, (b) decorrelation scale, and (c) believable scale. NRMSE is multiplied with 100 to facilitate comparisons. NRMSE has no unit, while decorrelation and believable scale are measured in grid boxes.

where the believable scale is expressed in [km], which facilitates comparisons to the decorrelation scale and eases interpretation. In contrast, the NRMSE is a number and is subject to double penalties.

Figure 2.8 illustrates the procedure in the idealized testbed. The observation is a Gaussian centered in the periodic domain with a total diameter of 20 grid boxes to represent a convective cell. It is unchanged for all tests. The forecast is identical, but will be modified in three ways to test the sensitivities of the metrics: (i) the effect of a displacement of the forecast (indicated by the arrows), (ii) the result of a too peaked or too wide forecast, and (iii) the effect of a too strong or too weak forecast.

Figure 2.9 shows a comparison of the behavior of NRMSE, decorrelation scale, and believable scale. We computed the scores for deterministic forecasts, which predict the correct Gaussian structure, but miss the exact location. The rapid loss in skill depicted by the NRMSE is known as the double penalty, which means that a small displacement rapidly causes large errors or even saturation of the score as not only is part of the precipitation missed, but also is another part falsely predicted. In contrast, both decorrelation and believable scale show smoother gradients and similar values. The progression of the believable scale is continuous, while steps are visible in the decorrelation scale. Those are a result of the spectral resolution of the FFT or DCT. As the spectral resolution for smaller scales, which will be more relevant in our study, is much higher, this issue is of minor relevance.

In Fig. 2.10a, the effect of a forecast with an incorrect size is visualized. In this situation, the NRMSE depicts a minimum for a correct forecast and increases smoothly and continuously for too small or too wide forecasts. Due to the shape of a Gaussian, the increase is not symmetric. This is also true for the decorrelation scale, although the growth is much slower, indicating that it penalizes a wrong-sized forecast less harshly. The believable scale rapidly grows from a nearly perfect forecast to the domain size around a size ratio of the forecast to the observation of 2. At this point, false positives outside of the compared neighborhood occur and impact the scale. As the FSS is a

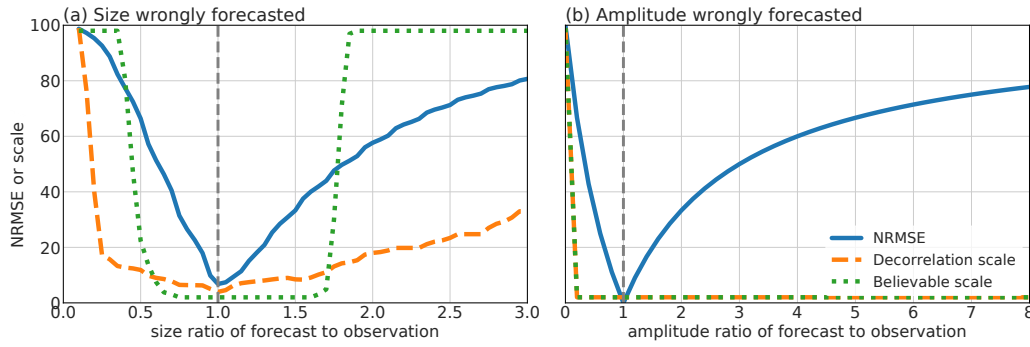


Figure 2.10: Verification of precipitation forecasts and the impact of size (a) and amplitude (b) of the forecast with different metrics: NRMSE, decorrelation scale and believable scale. NRMSE is multiplied with 100 to facilitate comparisons. NRMSE has no unit, while decorrelation and believable scale are measured in grid boxes.

domain-averaged score of grid point FSS values in this idealized case, the total domain size is of relevance.

Finally, we visualized the impact of a forecast with the wrong amplitude in Fig. 2.10b. We see that the amplitude has no implications on decorrelation and believable scale, and the forecast is recognized as *perfect*. In the case of the believable scale, this only holds as long as the threshold of the FSS is chosen in a way the fractions are not affected. The NRMSE is perfect for a ratio of 1 and increases for both too small and too large amplitudes in a continuous way.

We hope we could convince the reader of the complementary abilities of the applied metrics. Especially the fact that decorrelation and believable scale provide very similar results is of importance for the following discussion of our findings.

Predictability of Convective Precipitation

This chapter presents results on the practical predictability of convective precipitation in different model configurations ranging from highly idealized to realistic operational setups. For that purpose, we assess the impact of orography, radar DA, and their collective effect, as well as the consequences of a higher level of convective organization. Furthermore, we confirm and extend our findings in operational forecasts.

First, we focus on the effect of orography on the predictability in the presence of small-scale initial condition or small- and large-scale initial and boundary condition uncertainties in the idealized setup (see Sec. 3.1). Subsequently, we focus on the impact of radar DA and its potential interplay with the orography on the forecast in a perfect model environment (see Sec. 3.2). Section 3.3 investigates the influence of the convective mode or the level of organization on the predictability in the idealized system.

In a step towards a more realistic setting, we move to the pre-operational and operational NWP systems of DWD (see Sec. 3.4). First, we apply the introduced metrics to a two weeks period in summer 2016 that includes strong convection and features similar weather situations over Central Europe as the idealized experiments, but unlike the idealized setup, with a complex and realistic terrain (see Sec. 3.4.1). In this test period, we can also look into the impact on the performance of different DA schemes and the model's mixing length parameter, which is linked to convective triggering.

Finally, we verify our findings in a statistical sense for three summer periods from 2014 to 2016 in the then operational COSMO-DE-EPS system that relied on a combination of physics parameters and initial and boundary condition perturbations to initialize the ensemble (see Sec. 3.4.2). These experiments cover a wide variety of weather situations, which reintroduces another level of complexity and allows to consider the impact of synoptic-scale weather regimes.

As the present study focuses on predictability on convective scales, we applied elaborate neighborhood and spectral verification methods to deal with issues such as double penalties—the fact that a forecast of a slightly displaced convective cell might still be considered valuable by a forecaster. We think the application of identical metrics over the range of configurations is of seminal importance as it allows for transferability and fair comparisons.

3.1. Impact of Orography in Idealized Experiments

This section presents results from idealized experiments, starting with a description of the Nature Runs (IDEAL-Nat), with (*oro*) and without (*flat*) orography, in each case with two sets of ICs (IC and IBC) followed by the respective Background Ensembles (IDEAL). A comprehensive list of the abbreviations of all experiments including a short description is presented in Appendix 4.3 to for easier understanding.

We revisit the hypothesis by Anthes (1986) of increased convective predictability in the presence of orography in a convection-permitting EPS. The comparison of experiments with and without orography but otherwise identical meteorological conditions within the same model configuration allows for a clear attribution of all differences between the experiments to the orography. In the IDEAL IC experiments, we can study convective-scale predictability in a situation with only small-scale uncertainty and identical large-scale conditions, representing perfect synoptic-scale predictability. In contrast to IDEAL IC, the IDEAL IBC experiments introduce realistic synoptic-scale uncertainties to estimate their impact. Large sensitivities of precipitation patterns exist, according to Nuss and Miller (2001), in the vicinity of orography, due to small uncertainties in the synoptic flow. Large-scale uncertainties add realism to the experiments because every LAM inadvertently inherits those from an ultimately global model.

In the idealized experiments presented here, we strongly simplified the problem and performed numerical experiments with the operational COSMO system in an idealized setup. To account for the impact of orography on the predictability limits of convective precipitation, we excluded the following influences: (i) the uncertainty inherent in the weather situation by prescribing horizontally homogeneous atmospheric conditions with periodic boundaries, (ii) model errors by applying a perfect model approach, thereby neglecting approximations and inconsistencies in the description of atmospheric processes in the NWP model, and (iii) the interactions of complex terrain with the atmosphere by replacing it with a simpler Gaussian mountain.

3.1.1. Development of Convection in the Nature Run

Different stages of the convective life cycle of IDEAL-Nat *flat* and IDEAL-Nat *oro* are displayed in Fig. 3.1. At 0800 UTC (corresponding to 0 h lead time), where many convective cells with reflectivity values between 15 dBZ and 50 dBZ are present. In IDEAL-Nat *flat*, the cells appear randomly and spatially evenly distributed (see Fig. 3.1a,b), whereas in IDEAL-Nat *oro*, there is one strong convective cell (exceeding 45 dBZ) downstream of the orography (indicated by black circles) surrounded by predominantly weaker cells (less than 20 dBZ, see Fig. 3.1e,f).

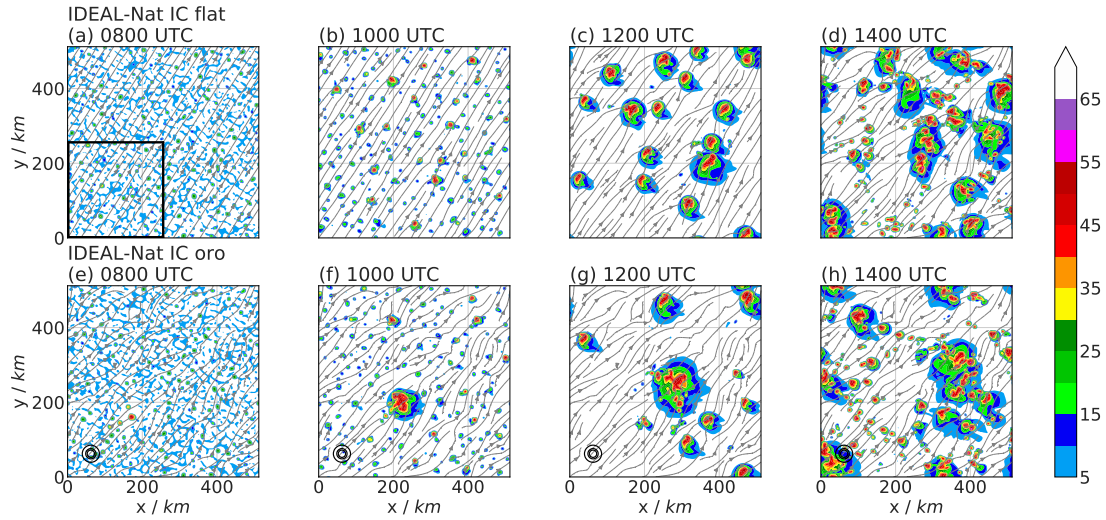


Figure 3.1: Maximum reflectivity of IDEAL-Nat *flat* (a–d) and IDEAL-Nat *oro* (e–h); the grey streamlines depict the horizontal wind at 1500 m above ground; plots are valid at 0800 UTC (a),(e), 1000 UTC (b),(f), 1200 UTC (c),(g) and 1400 UTC (d),(h); the black box in (a) visualizes the sub-domain referred to in the results section and the black circles in (e–h) to the location and height of the orography.

Larger convective cells and MCSs develop with progressing time due to high CAPE values and the persistent wind shear, while the spatial intermittency increases. The streamlines visualize the inflow and convergence regions of those systems at 1500 m. At this height, the average wind speed amounts to around 5 m s^{-1} . Growth and development of convection are also characteristic for individual members of IDEAL IC and, in a wider sense, the IDEAL IBC experiments, which we discuss in the next section.

As the *synthetic* observations for all idealized DA experiments in Sec. 3.2 are drawn from IDEAL-Nat (*oro* or *flat*), Fig. 3.1 represents a visualization of the reflectivity field prior to the processing in the radar forward operator. Note that IDEAL-Nat represent the *truth* for IDEAL IC and IDEAL IBC. As we want to compare those experiments, the respective Nature Runs for verification must be identical. Similarly, IDEAL-Nat represents the *truth* and provides the *synthetic* observations for the DA experiments IDEAL-DA IC and IDEAL-DA IBC.

3.1.2. Orography and Initial Conditions in the Ensemble Forecasts

In the following, we will discuss the evolution of the ensemble forecasts, which are based on the deterministic IDEAL-NAT experiments introduced above. The focus in this part lies on the effects of orography, and the combined effect of orography and large-scale uncertainties in the ICs on the predictability of convection in the idealized setup.

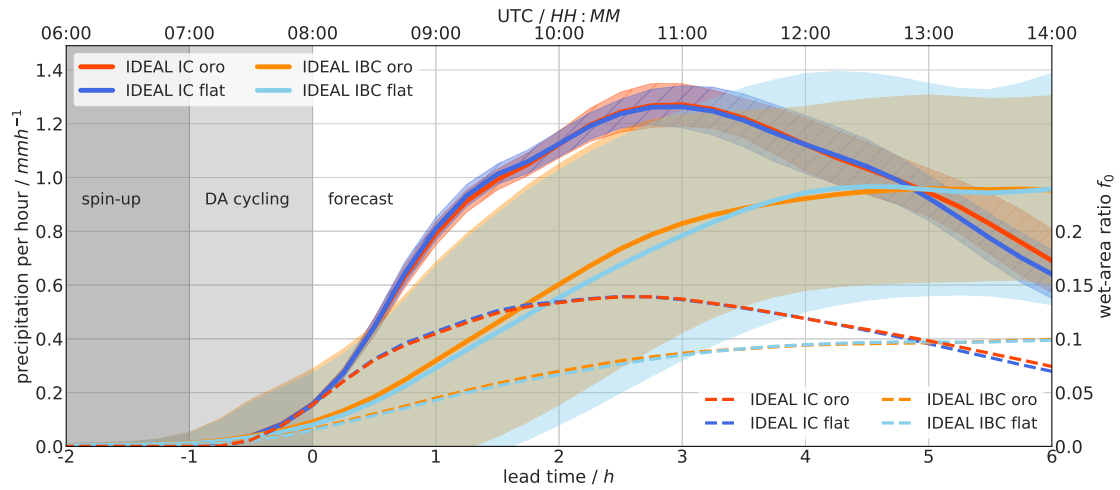


Figure 3.2: Time series of the domain-averaged hourly precipitation (solid lines) and wet-area ratio f_0 (dashed lines, threshold = 1 mm h^{-1}) for the IDEAL IC/IBC *oro/flat* experiments: the mean is indicated by thick lines, the standard deviation by shading, the spin-up period by dark grey and the window of the radar DA cycling by light grey. The hourly precipitation is indicated on the left and the wet-area ratio on the right.

As both IDEAL-Nat experiments are visually similar, the time series of domain-averaged hourly precipitation and the wet-area ratio only show negligible differences (see Fig. 3.2). There is hardly any precipitation in the spin-up period until 0700 UTC. Between 0700 and 0800 UTC (note that this is the DA window in later experiments), deep convection is initiated and precipitation reaches the ground. After 0800 UTC (0 h lead time), differences between the IDEAL IC and IDEAL IBC begin to show. In IDEAL IC, the precipitation strongly increases and peaks between 1000 – 1200 UTC (2 h – 4 h lead time), followed by a gradual decrease. The spread of IDEAL IC grows until 1200 UTC (4 h lead time) and is constant afterward. The wet-area ratio f_0 , i.e., the rainy fraction of the domain exceeding 1 mm h^{-1} , starts to grow shortly after 0700 UTC and peaks between 1000 – 1100 UTC (2 h – 3 h lead time) with values of 14 %. After that, f_0 gradually declines to 7 % at the end of the forecast (6 h lead time).

In contrast, IDEAL IBC depicts a slower growth in the precipitation rate until 1200 UTC (4 h lead time), and the amount of precipitation remains constant for the last 2 h of the forecast. The ensemble spread is drastically increased by the large-scale uncertainties introduced in the IDEAL IBC experiments, which also lead to the initiation of convection in a time period of 2 to 3 h. The wet-area ratio of IDEAL IBC follows a development similar to the one of the hourly precipitation. Precipitation rate and wet-area ratio are very similar for the pairs of IDEAL IC *oro/flat* and IDEAL IBC *oro/flat* throughout the forecasts, facilitating comparisons.

A snapshot of the spatial distribution of convective precipitation in all 40 members of the four IDEAL configurations and their respective IDEAL-NAT, which is identical for the pairs of IC/IBC experiments, is given in Fig. 3.3 for the 19 dBZ contour of the maximum reflectivity in the vertical column. The maximum reflectivity in the vertical column provides a good overview of the structure and location of the convection.

IDEAL: 0h lead time / 0800 UTC

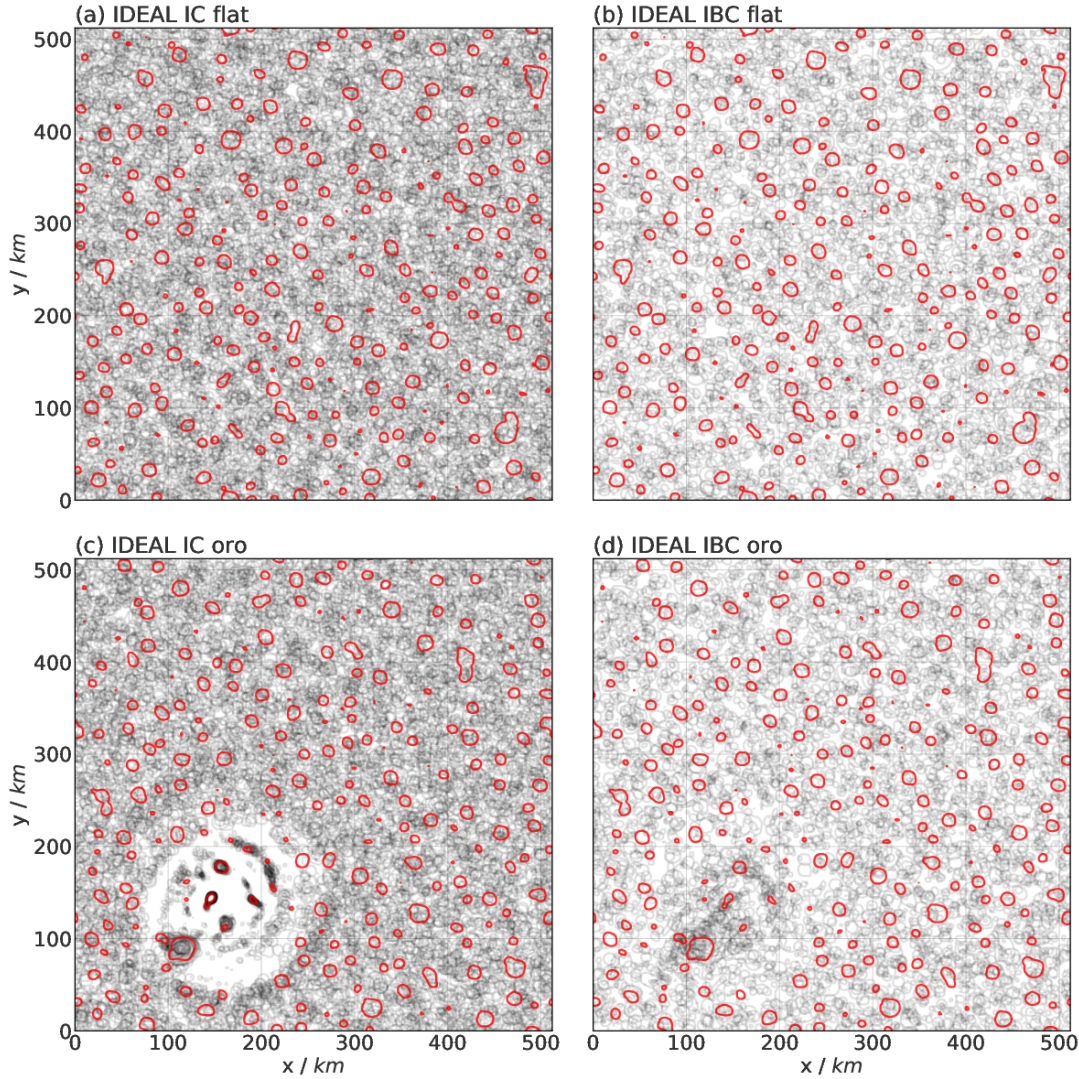


Figure 3.3: Spaghetti plots of the IDEAL ensemble at 0800 UTC / 0h: isolines of 19 dBZ of all 40 ensemble members (black) and the respective deterministic IDEAL-Nat (red). (a,b) depict the *flat* domain, (c,d) depict the *oro* domain; (a,c) with IC uncertainty and (b,d) with IBC uncertainty.

The so-called Spaghetti plots show the *truth* (IDEAL-NAT) in red and all 40 ensemble members (IDEAL) in black. Note that a reflectivity of 19 dBZ corresponds roughly to a precipitation rate of 1 mm h^{-1} and both fields lead to similar results for a variety of scores (Stratman et al., 2013). We choose this low threshold as ensemble forecasts typically do not show good reliability or resolution for rare events like strong convection, a tendency described by, for example, Murphy (1993) and Schwartz and Sobash (2017).

We will first describe the effect of orography in IDEAL IC *oro* and IDEAL IC *flat* before we will discuss the differences to IDEAL IBC. In an early phase of the convective life cycle, only small cells have developed (0800 UTC, 0h lead time, Fig. 3.3a). Since we apply an identical, horizontally homogeneous mean state to all ensemble mem-

IDEAL: 2h lead time / 1000 UTC

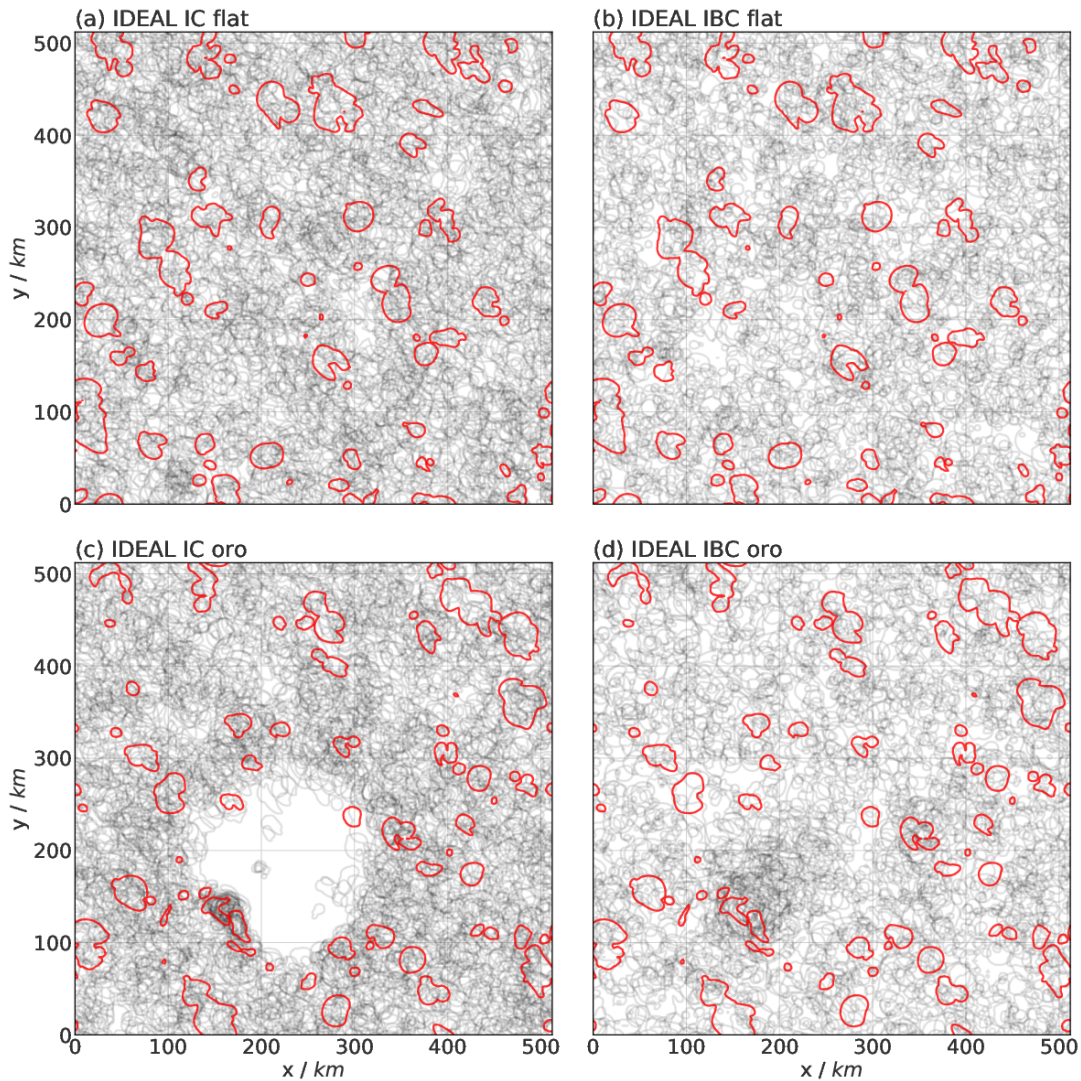


Figure 3.4: Spaghetti plots of the IDEAL ensemble at 1000 UTC / 2h: isolines of 19 dBZ of all 40 ensemble members (black) and the respective deterministic IDEAL-Nat (red). (a,b) depict the *flat* domain, (c,d) depict the *oro* domain; (a,c) with IC uncertainty and (b,d) with IBC uncertainty.

bers at the initial time, the cells' location is only determined by the initial noise, and the conditions are equally favorable for convection throughout the entire domain in IDEAL *flat*. Figure 3.3 already indicates that there is no predictability at the grid scale in IDEAL IC *flat*. In contrast, the orography causes convergence in its lee, which acts as a trigger for convection in IDEAL IC *oro*. This results in strong convective cells, restricted to a small area with favorable conditions for convection (see. Fig. 3.3c). These cells inhibit further convection in their vicinity and cause the 'dry patches' that see no precipitation in any of the ensemble members.

Demko and Geerts (2010) describe a similar situation close to an isolated mountain with comparable dimensions as in the present study (15 km radius and 2000 m height, in the

IDEAL: 4h lead time / 1200 UTC

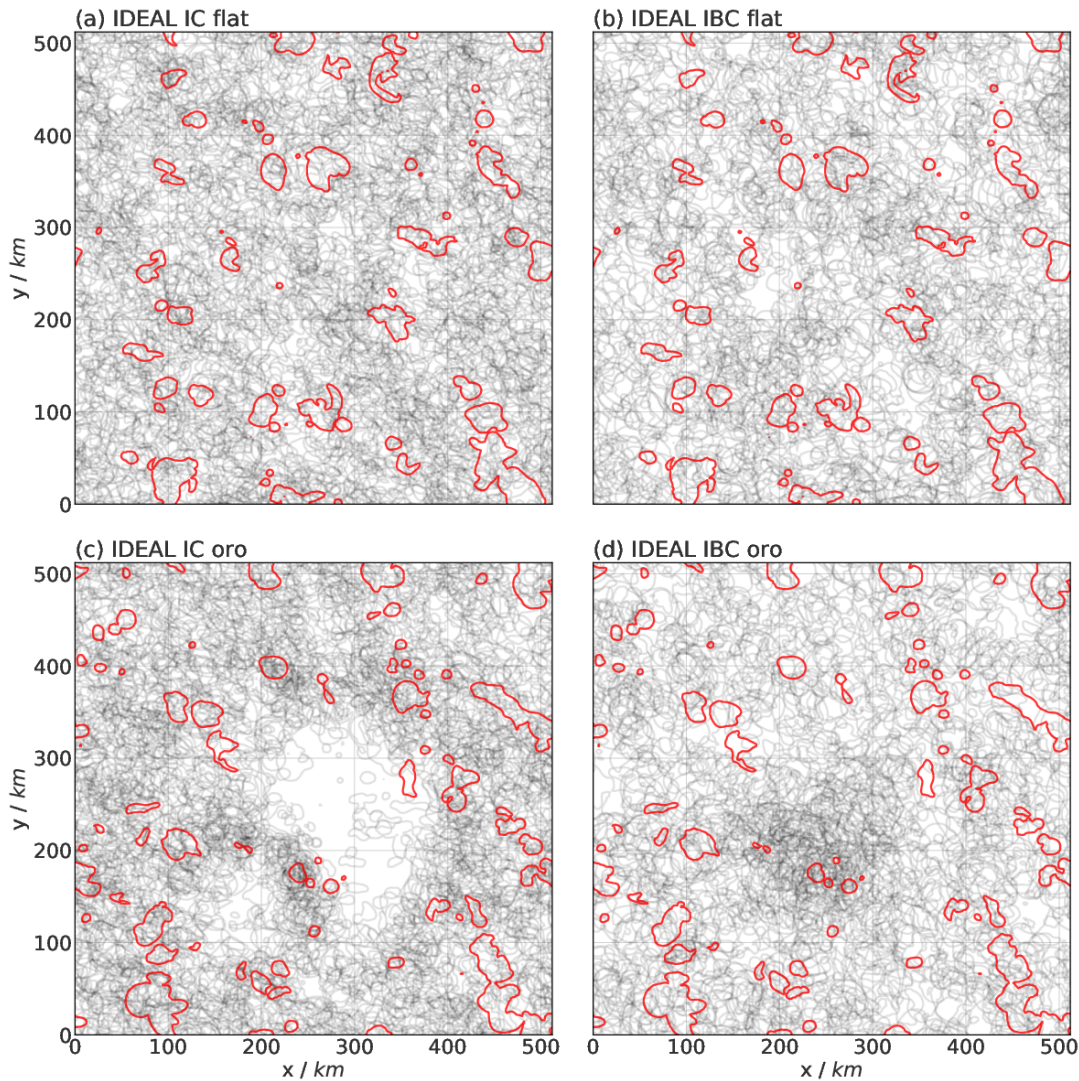


Figure 3.5: Spaghetti plots of the IDEAL ensemble at 1200 UTC / 4 h: isolines of 19 dBZ of all 40 ensemble members (black) and the respective deterministic IDEAL-Nat (red). (a,b) depict the *flat* domain, (c,d) depict the *oro* domain; (a,c) with IC uncertainty and (b,d) with IBC uncertainty.

Santa Catalina Mountains, Arizona) and also observe a dry region downstream of the orography. In idealized studies simulating various mountain heights, Hassanzadeh et al. (2014) and Panosetti et al. (2016) conclude that the exact locations of the convective cells and the dry regions depend on the specific environment and flow situation.

In comparison to IDEAL IC, both IDEAL IBC ensembles show higher variability in the location and timing of convection. In IDEAL IBC *flat* the position of convective cells in the ensemble is as random as in IDEAL IC *flat*, but the timing of convective initiation is also more variable. In combination with Fig. 3.2, the lighter grey indicates that not all ensemble members develop wide-spread convection at this time. IDEAL IBC *oro* shows similar characteristics as IDEAL IC *oro*, as there is still an area of increased convective

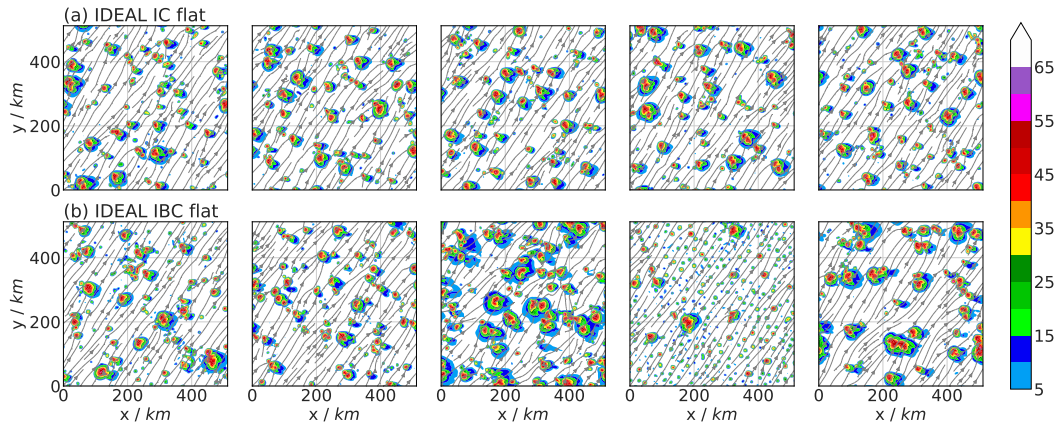


Figure 3.6: Horizontal cross-sections of five representative ensemble members of the IDEAL IC *flat* (a) and the IDEAL IBC *flat* experiments (b) at 1 h lead time (0900 UTC); the reflectivity is color-coded, the horizontal wind at 1500 m is indicated by grey streamlines.

activity and a relatively dry region, but both are more widespread and smeared out.

At a later stage of the convective life cycle, the number of cells has declined in favor of fewer but stronger convective systems, which are still fairly evenly distributed throughout the domain in IDEAL IC *flat* and IDEAL IBC *flat* at 1000 UTC (2 h lead time, see Fig. 3.4a,b). In contrast, the fingerprint of orography is clearly recognizable in IDEAL IC *oro* and IDEAL IBC *oro*, although the ICs cause differences (see Fig. 3.4c,d). Most ensemble members forecast a convective cell at (180km, 120km) and a mostly dry region downstream in IDEAL IC *oro*. The higher variability in IDEAL IBC *oro* leads to a comparatively larger region of increased precipitation with a diameter of roughly 100 km. The orographic influence on the cross-sections of reflectivity is restricted to the vicinity of the orography at 0800 UTC. 2 h and especially 4 h later, and the orographically influenced regions are advected downstream with the mean flow and grow substantially while losing sharpness. This is interesting as it visualizes how predictability can be transported in the system and how certainties decrease with lead time, in case of convection within just a few hours.

The visual inspection of the four IDEAL ensembles (IC/IBC; *flat*/*oro*) highlights increased predictability of convective precipitation in the presence of orography for several hours lead time. Visually this is, only evident in the ensemble visualization, but not in the visualization of the deterministic IDEAL-Nat (compare Fig. 3.1 and Fig. 3.3/3.4/3.5). The orographic forcing changes the probability for the occurrence of convection in its surroundings, while both lower and higher probabilities increase predictability at a certain position. In line with Hassanzadeh et al. (2014) and Panosetti et al. (2016), we also see a complex dependence of the precipitation patterns to the subtleties of the atmospheric flow. Foresti et al. (2018) recently published maps of the Swiss Alps, derived from a 10-year radar-based analysis, depicting the flow-dependent precipitation amount. From those maps, the connection mentioned above between precipitation and orography also becomes apparent.

Furthermore, convection modifies its vicinity, for instance by decreasing CAPE or triggering secondary convection via cold pools, thereby influencing areas further away from

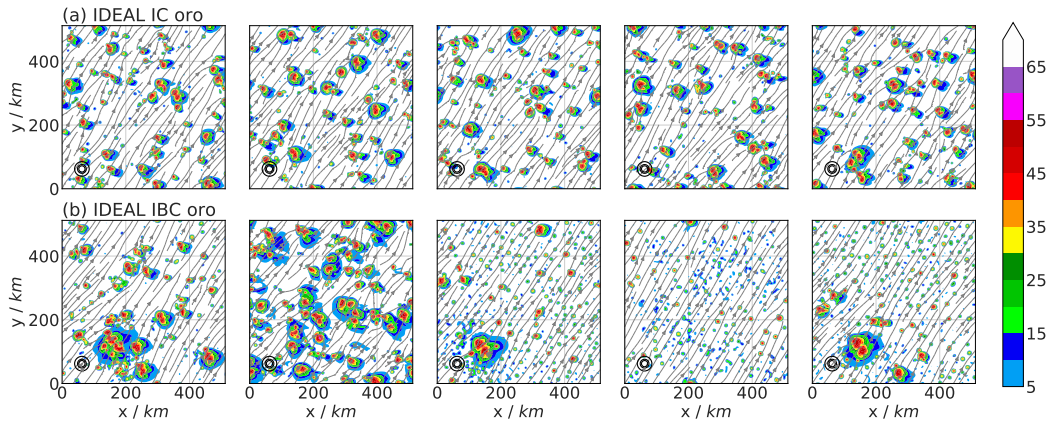


Figure 3.7: Horizontal cross-sections of five representative ensemble members of the IDEAL IC *oro* (a) and the IDEAL IBC *oro* experiments (b) at 1 h lead time (0900 UTC); the reflectivity is color-coded, the horizontal wind at 1500 m is indicated by grey streamlines and the black circles show the orography.

the orography as the altered environment is advected with the mean flow. As those processes take time, they can also increase the predictability for longer lead times and extend the forecast horizon, based on the visual impressions for several hours. The IDEAL IBC experiments show the effect of realistic, large-scale uncertainties in the ICs. A higher temporal and spatial variability is introduced that diminishes the beneficial impact of orography on the predictability, which nevertheless remains visible.

Figures 3.6 and 3.7 highlight another important aspect of the IDEAL IBC experiments. They visualize horizontal cross-sections of the column-maximum reflectivity of IDEAL *flat* (see Fig. 3.6) and IDEAL *oro* (see Fig. 3.7), valid at 0900 UTC (1 h lead time). The first row depicts five representative ensemble members started from only IC perturbations, while the second row features five members of IDEAL IBC with more realistic uncertainties.

In both sets of IDEAL IC members, convective cells of similar size throughout the domain can be seen, but the members differ on the exact location of the convection. In contrast, the sets of IDEAL IBC members show a range of convective cells of various sizes, and different convective scenarios can be seen. Those range from evenly distributed, predominantly small or large cells to members, where the convection is concentrated to a few strong cells. In the presence of orography, there is still a tendency of precipitation in the lee of the mountain (grey streamlines show the South-West wind). As did Nuss and Miller (2001), we observe a sensitivity of the precipitation patterns to the large-scale flow; however, it is present in the *oro* and *flat* experiments. Additionally to the spatial variability, the IDEAL IBC experiments (*oro* and *flat*) show higher temporal variability, with the onset of convection shifted by up to 2 h. In conclusion, the additional LBC uncertainties add variability and realism to the ensemble forecasts, which is lacking in IDEAL IC.

In the next step, the described effects are quantified by calculating the up-scaled NRMSE. As previously seen in the visual inspection of the location of precipitation, the

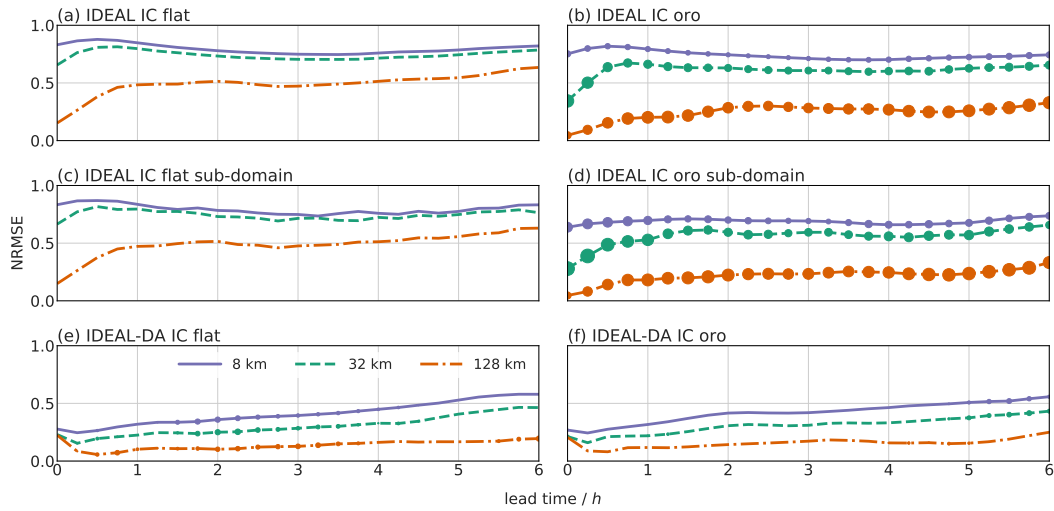


Figure 3.8: NRMSE of the column-maximum reflectivity for three levels of up-scaling (8, 32, 128 km); *flat* experiments are shown in the left column (a,c,e) and *oro* in the right (b,d,f); the first row (a,b) depicts IDEAL IC, the second row (c,d) a sub-domain around the orography and the third row (e,f) IDEAL-DA IC experiments; the dots highlight where *flat* outperforms the respective *oro* experiment in (a,c,e) and vice versa in (b,d,f), while their size is proportional to the difference in NRMSE.

NRMSE confirms that there is little predictive skill on scales of 8 km (see Fig. 3.8a,b), but the skill is increasing with a growing verification scale. In the experiment with orography, increased predictability can be seen on all scales. The size of the round markers indicates by how much IDEAL IC *oro* outperforms IDEAL IC *flat* and vice versa. Especially on the larger scale of 32 and 128 km IDEAL IC *oro* depicts consistently higher skill, which indicates that orography decreases displacement errors and accordingly increases the predictability of convective precipitation.

To highlight the impact of orography on precipitation in its vicinity, we focus on a sub-domain around the orography, i.e., the south-west quadrant encompassing 128×128 grid points of the domain (cf. the box in Fig. 3.1a), where the orography has visually the largest effect on the precipitation (see Fig. 3.3/3.4c). IDEAL IC *oro* again consistently outperforms IDEAL IC *flat* on all scales (compare dots in Fig. 3.8d). The strongest improvement in the sub-domain occurs on scales of 8 and 32 km and within the first 2 h. These scales represent typical sizes of developing and well-developed convective cells in the experiments and are therefore sensitive to displacement errors of these cells. In this period, many initially small convective cells develop and grow throughout the domain. In the sub-domain, the signature of the mountain is dominant, and the smoothing effect of averaging over the entire model domain is excluded. Note that the NRMSE of IDEAL IC *flat* over both domains exhibits no significant differences as the convection is evenly distributed throughout the domain (see Fig. 3.8a,c).

Figure 3.9a,b shows the NRMSE for the IDEAL IBC experiments. The feature of increasing skill for larger up-scaling values is reproduced (see Fig. 3.8a,b), while the overall skill in IDEAL IBC *flat* is slightly deteriorated. IDEAL IBC *oro* exhibits distinctly

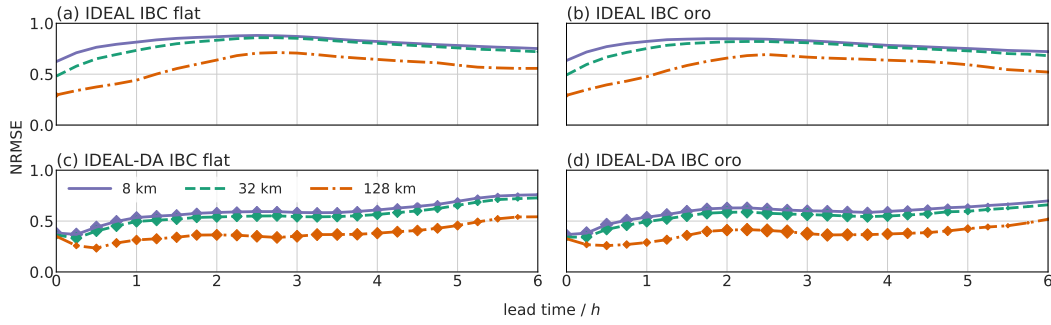


Figure 3.9: NRMSE of the column maximum reflectivity for three levels of up-scaling (8, 32, 128 km); *flat* experiments are shown in the left column (a,c) and *oro* in the right (b,d); the first row (a,b) depicts IDEAL IBC and the second row (c,d) IDEAL-DA IBC experiments; the diamonds highlight where IDEAL-DA IBC outperforms the respective IDEAL IBC experiment in (a,b) and vice versa in (c,d), whereby their size is proportional to the difference in NRMSE.

worse skill on all scales compared to the simulations without LBC uncertainties and the differences between IDEAL IBC *flat*/*oro* vanish. The results suggest that the realistic large-scale uncertainty can impair the beneficial impact of orography mentioned before to some degree.

Furthermore, the decorrelation scale, which requires no selection of a threshold and which is calculated in spectral space, is displayed in Fig. 3.11a. Since IDEAL IC/IBC *flat* are initialized from ICs based on random, uncorrelated noise, the position of the convective cells is uncorrelated and the decorrelation scale equals the domain size of 512 km throughout the experiments. In contrast, the decorrelation scale of IDEAL IC *oro* is much smaller, starting with around 10 km and increasing to 70 km within the first 3 h while saturating thereafter. The impact of additional large-scale uncertainties on the predictability of deep convection is investigated comparing the IDEAL IC with IBC experiments. IDEAL IBC *oro* depicts, with a decorrelation scale of 40 km at 0 h lead time, considerably higher values than IDEAL IC *oro*. This difference persists throughout the forecast and amounts to 50 km after 6 h (50 km IDEAL IC *oro* and 100 km IDEAL IBC *oro*).

As both IDEAL *oro* experiments are initialized in the same way as IDEAL *flat*, the scale reduction we observe can be attributed to the presence of orography. Comparing IDEAL IC *oro* and IDEAL IBC *oro* shows the detrimental impact of large-scale uncertainty that is in reality inherited from an (ultimately global) model providing the ICs and LBCs. It also highlights the potentially profound positive impact that improvements in the global model that provides those ICs can have. Overall, the decorrelation scale consolidates our visual impressions.

Before we discuss the believable scale, which is based on the FSS (compare Sec. 2.5.4), we want to show the original FSS for a threshold of 19 dBZ. In contrast to the decorrelation scale, the believable scale includes a comparison to, in this case, *synthetic* observations. In Figure 3.10, FSS tables with color-coded values of the FSS for the entire lead time of 6 h, with temporal resolution 15 min and spatial resolution 2 km, are

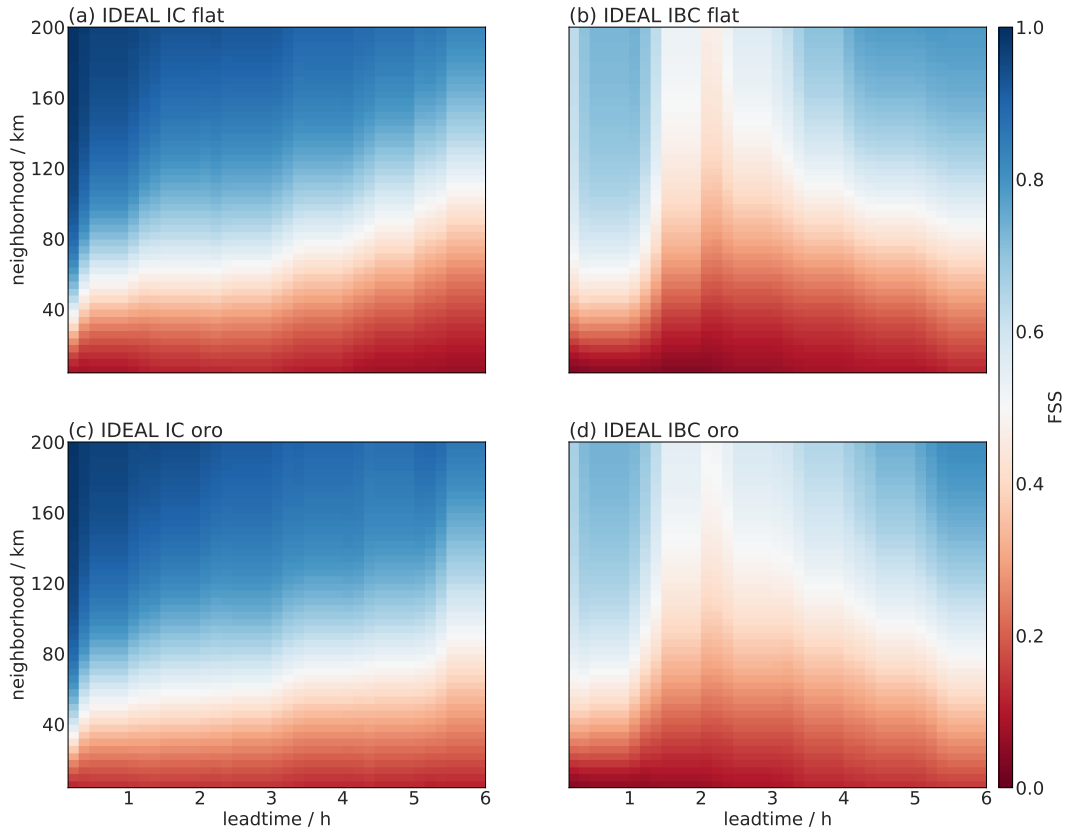


Figure 3.10: FSS tables of the IDEAL ensembles for a threshold of 19 dBZ: the color-coding represents the mean FSS score of all 40 ensemble members. (a,b) depict the *flat* domain, (c,d) depict the *oro* domain; (a,c) with IC uncertainty and (b,d) with IBC uncertainty.

shown. Blue indicates predictable scales and red unpredictable ones, while the transition zone in white is similar to the believable scale, which is defined as the scale with skill $0.5 + f_0$. By accounting for the wet-area ratio f_0 , the believable scale corrects for the positive effect of larger precipitation coverage on the score. However, the complete FSS provides additional information, for example, how skillful forecasts are at a certain time and scale.

The comparison of the IDEAL experiments reveals that IDEAL-IC *oro* and IDEAL-IC *flat* perform comparably good, and that both experiments reach high skill for short lead times and scales above 80 km. Similarly, both IDEAL-IBC experiments (*oro* and *flat*) depict a comparable skill over all scales and lead times. However, the large-scale uncertainty leads to significantly worse forecasts, especially after 1 h lead time. In a forecast range of 3 – 6 h the forecasts seem to become even slightly more skillful. They also depict a smaller gradient around a value of 0.5, due to a larger variability in the ensemble members causing larger variability in the FSS of the member.

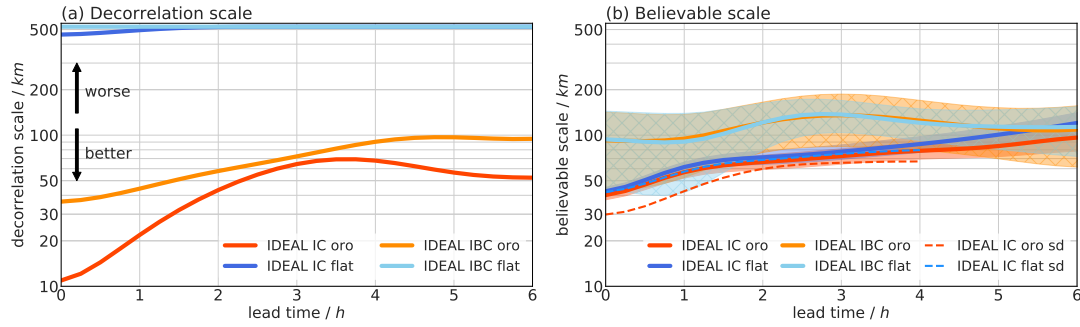


Figure 3.11: Time series of the decorrelation scale of the IDEAL experiments (a) and the believable scale for a threshold of 19 dBZ of the IDEAL experiments and a sub-domain in the vicinity of the orography (b).

In order to substantiate and extend the insights gained by the FSS, we provide the believable scale also calculated for a threshold of 19 dBZ of the column-maximum reflectivity (see Fig. 3.11b). It shows a gradual increase in the believable scale in IDEAL IC *flat* and *oro* with time. Initially, the believable scale amounts to 40 km and increases to 60 km within the first forecast hour. The values of both experiments remain similar for about 3 h. Thereafter, IDEAL IC *oro* performs better. After 6 h, the believable scale reaches approximately 100 km in IDEAL IC *oro* and 120 km in IDEAL IC *flat*, highlighting the skill improvement introduced by orography. The largest difference can be seen in later forecast hours (after more than 3 h), due to the continuous orographic forcing on the flow and propagation of the influence by downstream advection. At this point in time, the standard deviation of the ensemble members is still below 10 km (20 km after 6 h). Differences between the FSS and believable scale are expected and explained by the correction for the precipitation coverage.

The effect of the large-scale uncertainties on the believable scale can also be seen in Fig. 3.11b and is striking. Both IDEAL IBC experiments (*oro* and *flat*) are, as expected, degraded in comparison to the IDEAL IC reference. Their believable scale remains between around 100 - 125 km for the entire lead time with a standard deviation of the ensemble members of 50 km. The fact that the mean and the standard deviation exhibit no trend over time indicates that the believable scale saturates for the displacement of convection in the simulated situation. This value compares to a climatology of the situation set by the average conditions for deep convection in the domain, a finding in line with the small trend of the decorrelation scale of IDEAL IBC *oro* in Fig. 3.11a. Additionally, the boundary condition perturbations seem to overwhelm the positive impact of the orography on the predictability measured by the believable scale (comp. IDEAL IC/IDEAL IBC in Fig. 3.11).

For the believable scale, we additionally focused on the sub-domain adjacent to the mountain. In the case of the decorrelation scale, which is calculated in spectral space, this would render any result obtained hard to interpret. In the first two hours, the believable scale of IDEAL IC *oro* is roughly 20 km smaller than in IDEAL IC *flat* in that sub-domain. Note that the values of IDEAL IC *flat* remain similar to those of the whole domain, indicating that the sub-domain still represents a statistically relevant sample.

We conclude that the impact of orography possesses a certain range depending on the height and structure of itself, as well as the conditions of the atmospheric flow.

In the absence of model errors in this idealized setup, the decorrelation and believable scale highlight similar features, as the predictability of the model and atmospheric state are the same. The values of the decorrelation scale are somewhat smaller than the believable scale, but still in the same range. Both measures evaluate forecasts in physical and spectral space, respectively, and hence no one-to-one equivalent can be expected in a realistic and complex situation. In summary, orography leads to a noticeable increase in the practical predictability of convective precipitation in idealized experiments. The influence is most substantial in the close vicinity of the mountain and on the scale of well-developed cells (8 and 32 km, see Fig. 3.8), indicating increased predictability concerning the position of convection.

We find that large-scale uncertainties add realism to the idealized experiments and provide us with additional insights. The large-scale uncertainties seem to overwhelm the positive impact of orography on the predictability (compare IDEAL IC *oro*/IDEAL IC *flat* with IDEAL IBC *oro*/IDEAL IBC *flat* in Fig. 3.11). A finding in line with Nuss and Miller (2001) and Picard and Mass (2017), who found orographically influenced local precipitation patterns to be sensitive to perturbations of the large-scale flow direction.

3.2. Impact of Radar Data Assimilation in Idealized Experiments

We have seen that orography can increase the predictability of convection, but that a large portion of this improvement is lost in the presence of realistic large-scale uncertainties. Therefore, we are interested in how the impact of radar DA and orography compare to each other given that both exert an effect of the position of convective precipitation. Do the beneficial effect on the forecast quality amass if radar DA is implemented in the presence of orography? And is the assimilation of mesoscale features, such as convection, able to correct large-scale uncertainties?

As in Section 3.1, we strongly simplify the problem and perform numerical experiments with the operational COSMO-KENDA system in an idealized OSSE setup. In addition to the influences we excluded so far, namely variability in the weather situation, model errors and interactions between atmosphere and *complex* orography, this study can benefit from further constrictions on the DA. Therefore, we exclude (i) observation uncertainties as representativity issues or systematic and correlated errors by using a perfect model OSSE approach creating spatio-temporally highly resolved *synthetic* observations from a Nature Run, and (ii) thereby neglect approximations and inconsistencies in the description of atmospheric processes in the NWP model or the forward operator.

This section presents results from idealized OSSEs including radar DA, with (*oro*) and without (*flat*) orography, and in each case with two sets of ICs (IC and IBC). The performance and impact of radar DA in those experiments is discussed before we test sensitivities of the forecasts to changes in the DA configuration in a subset of the abovementioned experiments. The abbreviations for the presented experiments, including those DA configurations, and their meaning are summarized in Table 4.3 in the Appendix.

3.2.1. Orography and Initial Conditions under Radar DA

The impact of radar DA on the predictability of precipitation, as well as the quantification of the predictable scale as a function of lead time, is investigated by applying the same hierarchy of evaluation methods as before. We begin with the ensemble experiment IDEAL-DA IC, which is initialized from an analysis provided by assimilation of radar reflectivity and velocity information (see Fig. 2.3). IDEAL-DA IC can be regarded as a benchmark experiment showing the potential of radar DA in a favorable situation, i.e., DA during an early phase of the convective life cycle (here 0700 – 0800 UTC) using high-quality observations (derived from IDEAL-Nat with comparably small and uncorrelated observation errors) and a well-forecasted large-scale situation.

The visual comparison of the spatial distribution of convective precipitation in the 40 member IDEAL-DA IC experiments at 0800 UTC (*flat* and *oro*) with the IDEAL IC experiments (compare Fig. 3.3/3.12) already demonstrates the large beneficial impact of radar DA. The LETKF creates or suppresses convective cells by changing the associated weights of the ensemble members depending on their agreement with the radar observations. Therefore, ensemble members of IDEAL-DA IC agree to a higher degree with each other and the IDEAL-Nat on the cells' positions. However, some false positives and misses of detection are still present. Downstream of the orography, where agreement within IDEAL IC *oro* was already high, radar DA further improves the visual agreement in IDEAL-DA IC *oro* (see Fig. 3.12).

The visual differences between IDEAL-DA IC and IDEAL-DA IBC are less apparent after DA is applied, but still visible at 0800 UTC. In IDEAL-DA IBC (*flat* and *oro*), the agreement among the ensemble members is diminished and more substantial discrepancies to IDEAL-Nat show. The precise distribution of precipitation and dry patches that we see in IDEAL-DA IC *oro* is visible, but fuzzier in IDEAL-DA IBC *oro*. However, the excellent agreement at the analysis time shows that KENDA is able to converge towards the truth in a convective situation and in the presence of large-scale uncertainty.

We think it is worth mentioning at this point that the DA is not aiming at producing the same analysis for IDEAL-DA IBC as for IDEAL-DA IC. KENDA is a version of an EnKF and therefore uses a flow-dependent model error covariance matrix \mathbf{P} estimated with the ensemble, which accounts for the uncertainty of the weather situation. As the first guess of IDEAL-DA IBC shows higher variability than IDEAL-DA IC (compare Fig. 3.2, 3.3), while the observation errors are identical, the relative weight of the observations is increased and, therefore, the analysis increments applied to the ensemble mean become on average larger. However, the *theoretical* analysis spread is different, as the situation in IDEAL-DA IBC is more uncertain.

Figures 3.13 and 3.14 depict the same experiments as Fig. 3.12, but 2 h (4 h) later at 1000 UTC (1200 UTC). First, the on average larger cells highlight the evolution of the convection in all IDEAL-DA experiments. Secondly, the agreements between all IDEAL-DA experiments themselves and the respective IDEAL-Nat are still very high after 2 h lead time (see Fig. 3.13), and even after 4 h forecast lead time, we can see good agreement in IDEAL-DA IC, while the convection in IDEAL-DA IBC starts to diverge spatially. Thirdly, the additional large-scale uncertainty in IDEAL-DA IBC leads to more variability in the location and size of the convection and an increasing disagreement with the respective Nature Run with progressing lead time.

Application of the NRMSE supports these impressions for IDEAL-DA IC. Both IDEAL-DA IC experiments (*flat* and *oro*) exhibit a large reduction in NRMSE compared to IDEAL IC for all lead times and scales (see Fig. 3.8e,f). Radar DA further improves the up-scaled NRMSE at the scale of 128 km, which was already skillful in the IDEAL IC experiments. However, its positive effect is more pronounced on scales of 8 and 32 km. Those scales experience the largest improvements as we assimilate observations of a mesoscale phenomenon, leading to the largest correlations in the DA at those scales. The constant horizontal localization of 32 km, limiting the horizontal reach of the observations, enhances the effect.

IDEAL-DA: 0h lead time / 0800 UTC

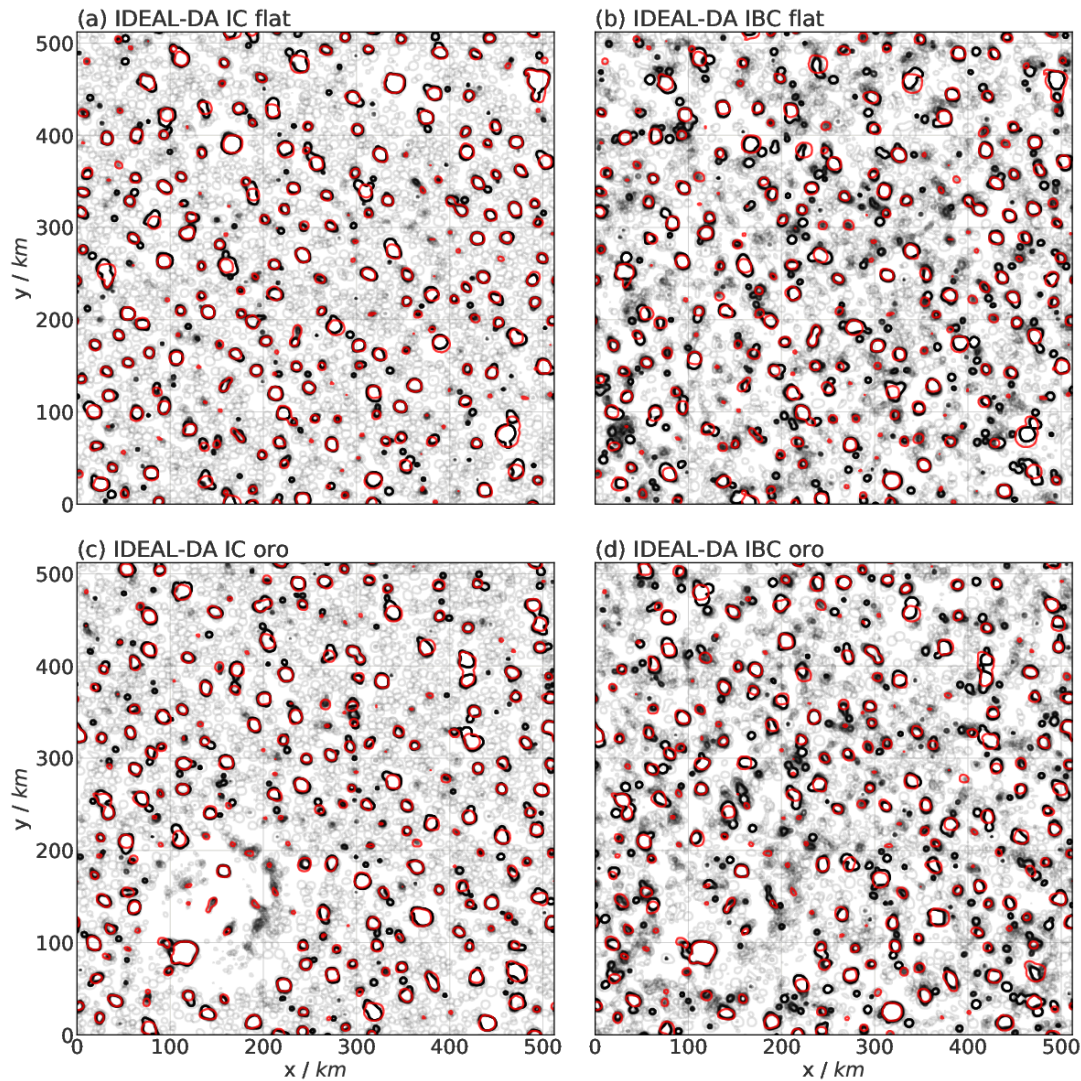


Figure 3.12: Spaghetti plots of the IDEAL-DA ensemble at 0800 UTC / 0 h: isolines of 19 dBZ of all 40 ensemble members (black) and the respective deterministic IDEAL-Nat (red). (a,b) depict the *flat* domain, (c,d) depict the *oro* domain; (a,c) with IC uncertainty and (b,d) with IBC uncertainty.

After radar DA, the NRMSE values of IDEAL-DA *flat* and IDEAL-DA *oro* amount to less than 0.6 at a lead time of 6 h, showing the beneficial impact of radar DA throughout the entire short-range forecast. The effect of radar DA in the IDEAL IBC experiments can be seen in Fig. 3.9. The skill in the IDEAL-DA IBC experiments is not as good as in IDEAL-DA IC, but as in the IDEAL IC experiments, radar DA has a largely beneficial impact on all scales for around 5 h, and the differences between *oro* and *flat* are negligible. The improvement by DA is visualized by the size of the diamond marker in the figure.

The relative impact of orography is reduced using radar DA. While the beneficial impact of orography is visible in IDEAL throughout the 6 h lead time, the IDEAL-

IDEAL-DA: 2h lead time / 1000 UTC

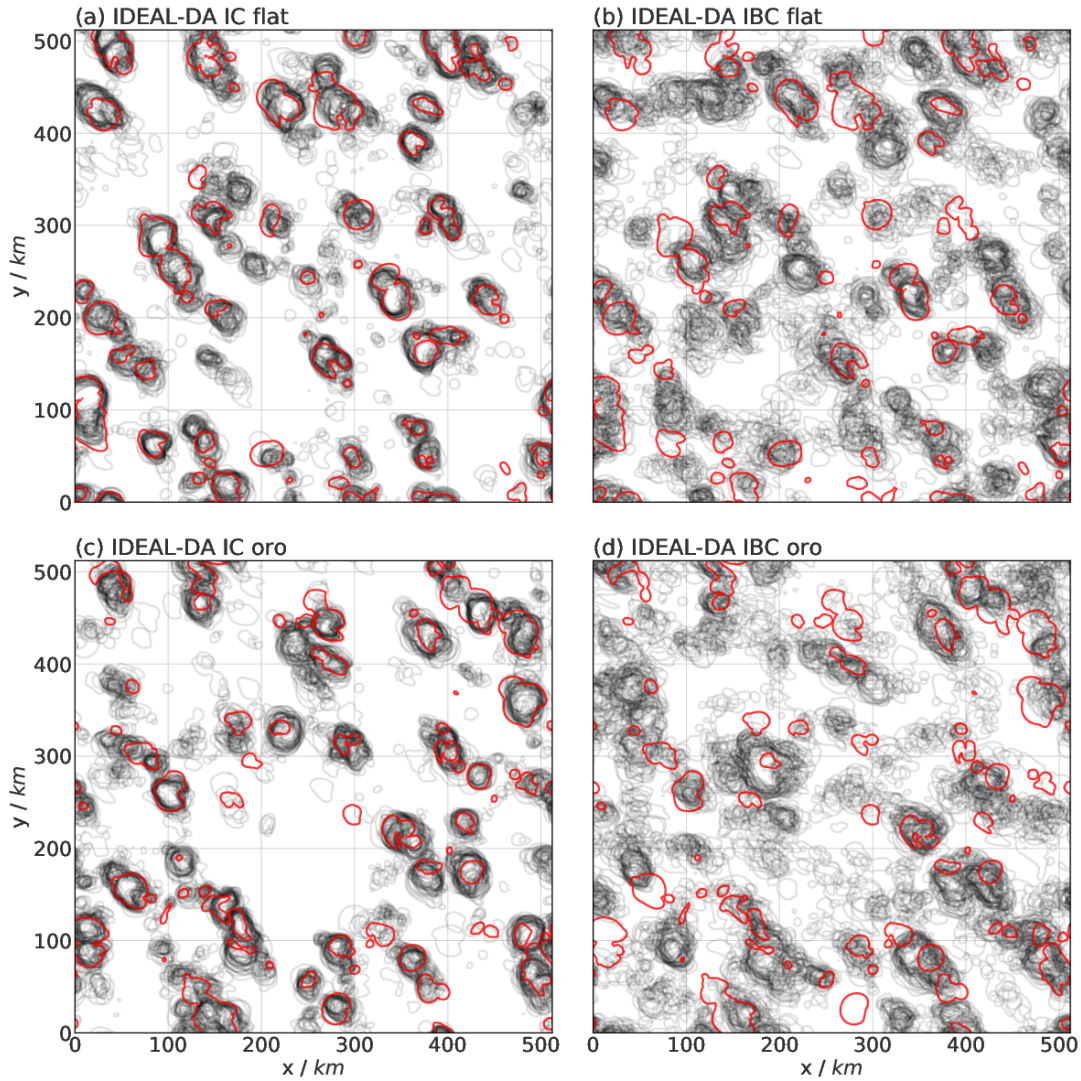


Figure 3.13: Spaghetti plots of the IDEAL-DA ensemble at 1000 UTC / 2 h: isolines of 19 dBZ of all 40 ensemble members (black) and the respective deterministic IDEAL-Nat (red). (a,b) depict the *flat* domain, (c,d) depict the *oro* domain; (a,c) with IC uncertainty and (b,d) with IBC uncertainty.

DA experiments exhibit comparable values of NRMSE with both sets of ICs (IC/IBC), which is consistent with results of Kühnlein et al. (2014) demonstrating that ICs have a dominating effect on the forecast error within the first 6 h before LBC and physics perturbations attain a comparable impact.

Interestingly, in all IDEAL-DA experiments (IC and IBC), *oro* and *flat* perform equally well in the decorrelation and believable scale, thereby highlighting that direct assimilation of high-quality radar observations can compensate the positive impact of orography on predictability. Therefore, we do not distinguish between *oro* and *flat* experiments in the explanations for the rest of this section. In the provided figures, however, the pairs of *oro/flat* experiments can be seen.

IDEAL-DA: 4h lead time / 1200 UTC

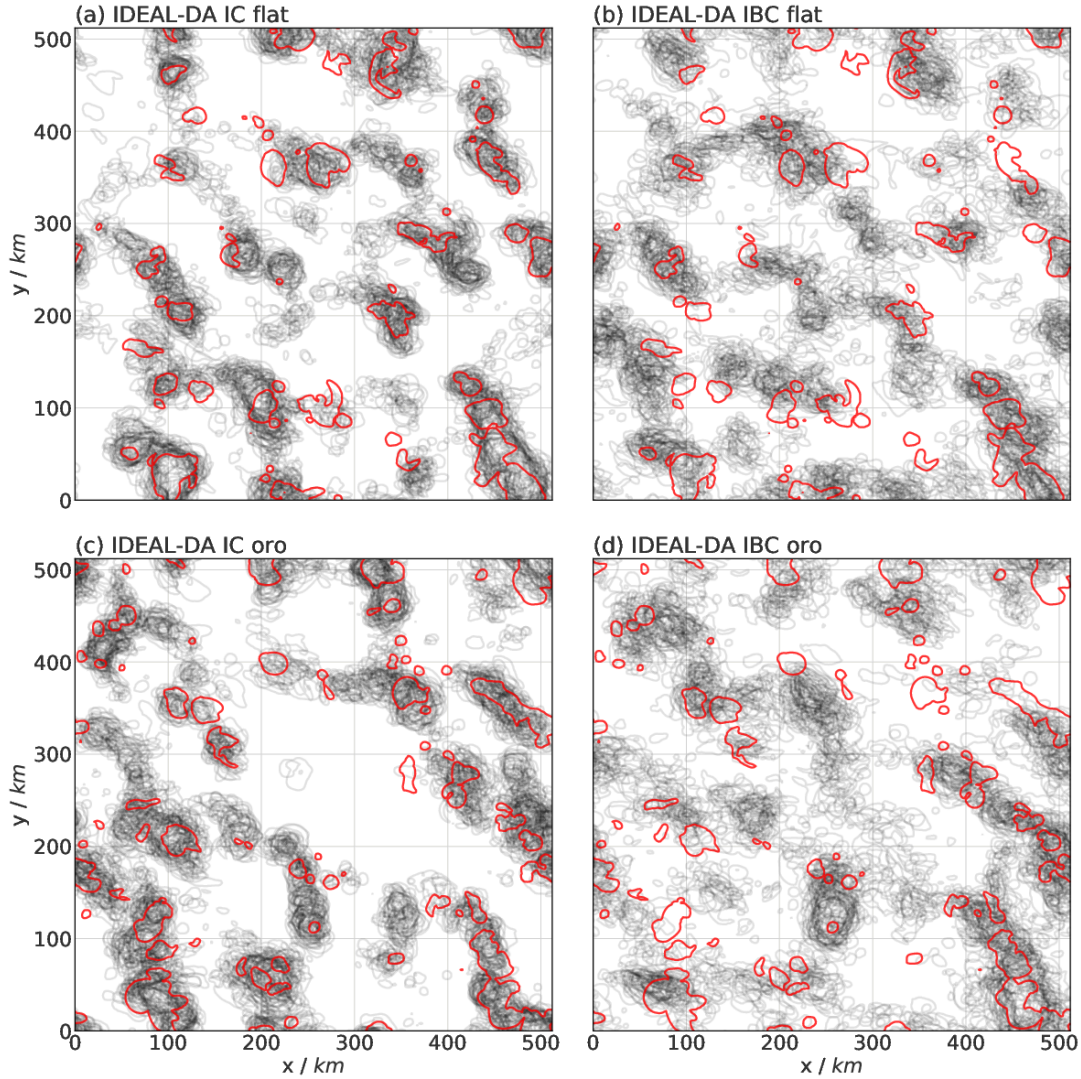


Figure 3.14: Spaghetti plots of the IDEAL-DA ensemble at 1200 UTC / 4 h: isolines of 19 dBZ of all 40 ensemble members (black) and the respective deterministic IDEAL-Nat (red). (a,b) depict the *flat* domain, (c,d) depict the *oro* domain; (a,c) with IC uncertainty and (b,d) with IBC uncertainty.

Radar DA shows a profoundly beneficial impact on the forecast skill in the idealized experiments with only IC uncertainties (IDEAL-DA IC). Measured by the decorrelation scale, the assimilation of *synthetic* radar observations increases the already high predictability further from 10 km to 4 km at 0 h lead time in IDEAL-DA IC *oro*. IDEAL IC *flat* exhibits no predictability of the location of convection without DA, but for IDEAL-DA IC *flat*, the decorrelation scale also drops to 4 km at 0 h lead time, which is the smallest resolved scale in spectral space. In the course of the simulations, the decorrelation scale gradually increases to about 30 km, which translates to an increased forecast horizon of several hours.

In the following, we investigate the performance of radar DA under impaired conditions

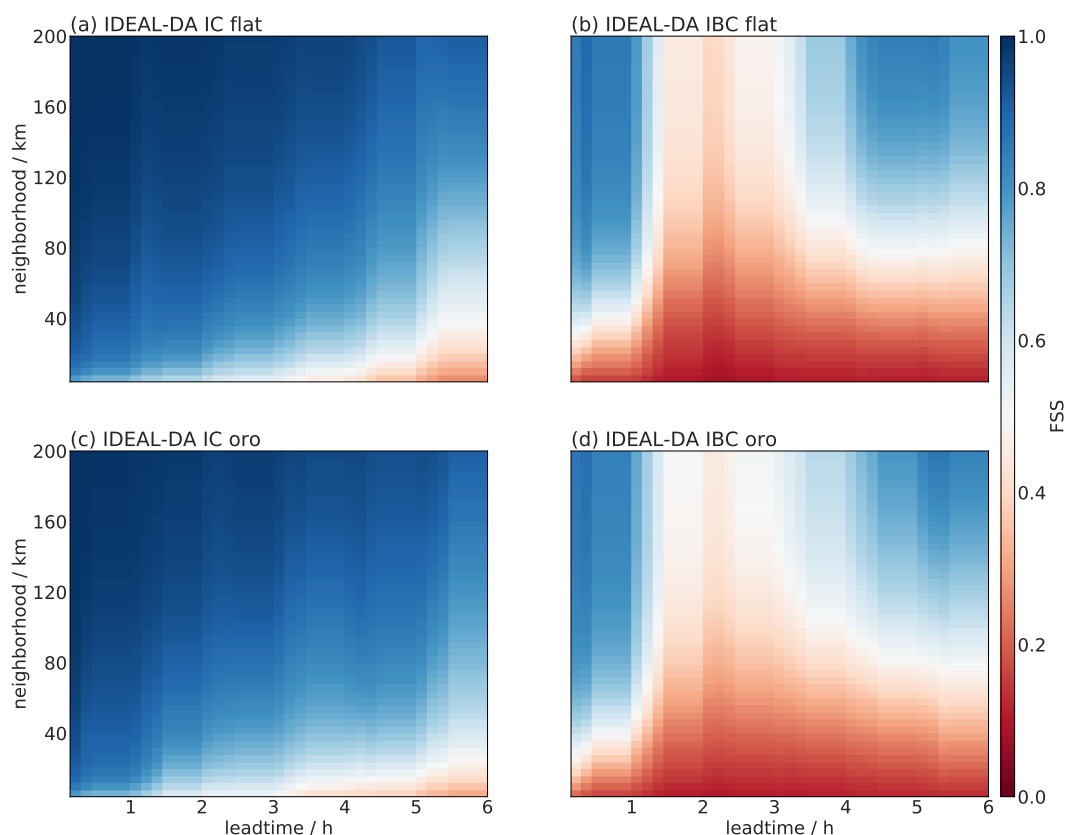


Figure 3.15: FSS tables of the IDEAL-DA ensembles for a threshold of 19 dBZ: the color-coding represents the mean FSS score of all 40 ensemble members. (a,b) depict the *flat* domain, (c,d) depict the *oro* domain; (a,c) with IC uncertainty and (b,d) with IBC uncertainty.

in the presence of large-scale uncertainty. The decorrelation scale of the smallest resolved scales in IDEAL-DA IBC becomes decorrelated and therefore unpredictable even at the initiation of the forecast. Nonetheless, the difference between IDEAL-DA IC and IDEAL-DA IBC stays below 10 km for the entire forecast, which translates to a forecast horizon shortened by 30 to 60 min. IDEAL-DA and IDEAL-DA IBC consistently and distinctly outperform the experiments without DA in terms of the forecast horizon by over 5 hours.

As in the previous section, we want to discuss the full FSS tables before the believable scale. It is visible that the effect of the initial conditions (IC/IBC) is substantially larger than the effect of orography (*oro/flat*) on the skill of the forecast. We see high predictability for scales as small as 40 km in the IDEAL-DA IC experiments, while predictability on these scales is lost after about 90 min in IDEAL-DA IBC. Also, the overall effect of radar DA in the presence of large-scale uncertainties seems rather small (compare Fig. 3.15 and 3.10). This is a case that highlights the usability of the believable scale, as it simplifies the information of the table to a concise line plot, easing

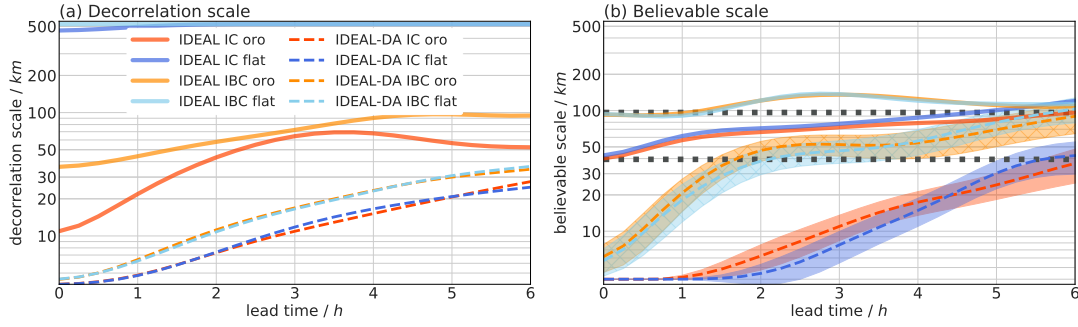


Figure 3.16: Time series of the decorrelation scale of the IDEAL/IDEAL-DA experiments (a) and the believable scale for a threshold of 19 dBZ (b). The increased forecast horizon due to radar DA is indicated by the dotted, horizontal lines.

comparisons drastically.

The believable scale depicts a significantly higher skill in the IDEAL-DA IC experiments than in IDEAL IC (Fig. 3.16b), while showing similar behavior throughout the forecast, starting with a believable scale of 2 km, the grid scale, at the initial time and increasing to 40 km after 6 h lead time.

The initial effect of radar DA on the believable scale is even larger in IDEAL-DA IBC, with large-scale uncertainties. It improves the predictable scales from the abovementioned climatological level of $\mathcal{O}(100 \text{ km})$ to scales of $\mathcal{O}(10 \text{ km})$ at 0 h lead time and only reaches it again after 6 h lead time. Comparing the IDEAL-DA IC and IDEAL-DA IBC experiments, the believable scale is, as visualized in Fig. 3.16b, almost identical at the start of the forecast but grows faster and reaches values around 50 km after 90 min, a feature that was already visible in the FSS table (see Fig. 3.15).

IDEAL-DA IBC also shows a larger standard deviation within the ensemble, which was also visible in the Spaghetti plots (see Fig. 3.3). As explained there, the EnKF recognizes IDEAL-DA IBC as the more uncertain situation and adapts the ensemble spread accordingly. After 6 h lead time, IDEAL-DA IBC reaches values around 100 km, while IDEAL-DA IC still remains below 50 km. The variability of the believable scale is, due to the higher variability of the atmospheric state in IDEAL IBC, also increased in IDEAL-DA IBC compared to IDEAL-DA IC. Apart from the additional information on the variability, the believable scale shows similar behavior as the decorrelation scale.

A comparison of these findings with the results of Sec. 3.1 highlights a higher impact of radar DA on the predictability of convection than orography in both sets of ICs. After the data assimilation, the performance of experiments with orography is indistinguishable to those without orography, showing that the beneficial impacts of the assimilation of radar data and orography are not accumulative in the idealized environment, which might be due to the sizeable beneficial impact of the data assimilation.

We attribute the large beneficial impact of radar DA on the predictability of convection in our experiments to several factors. First, this study assumes a perfect forecast model, which excludes error sources such as unresolved up-drafts or parametrizations

of microphysics and turbulence, which are highly relevant for deep convection. Secondly, we assume excellent conditions for radar DA, without systematic model and observation operator errors and a radar with comparably low and uncorrelated observation errors. And finally, in IDEAL-DA IC, there is no large-scale IC and LBC error that an operational LAM inherits from a global model.

However, we accounted for the effect of large-scale uncertainty in the IBC experiments and, although the effect was significant in our experiments, the increased forecast horizon of 6 h is still visible in the IDEAL IBC experiments (horizontal lines in Fig. 3.16). Therefore, we interpret our results as a potential upper limit for the impact of radar DA on the practical predictability. The impact of our assumed, and advanced radar will be discussed in Section 3.2.2.

The comparison of IDEAL with IDEAL-DA underscores the beneficial impact of radar DA on the predictability of convection in COSMO-KENDA. Radar DA provides a promising tool to increase the practical predictability of deep convection on a time frame of several hours. The IDEAL-DA experiments outperform IDEAL by about 40 km with only IC perturbations and by up to about 80 km with additional LBC perturbations. The KENDA system proved to be effective in correcting the position of deep convection in both setups, especially in IDEAL-DA IC, and, in the idealized setup without model errors, COSMO propagates and develops the convection properly. As expected, additional LBC perturbations lead to reduced predictability, but still valuable forecasts provided radar DA is applied, still valuable forecasts. The comparable skill of IDEAL-DA *flat* and IDEAL-DA *oro* emphasizes that the inherently higher predictability in the presence of orography can be compensated by radar DA for several hours and that parts of the lost predictability due to large-scale uncertainties can be recovered by radar DA.

3.2.2. Sensitivities of the DA system

In this section, several sensitivities of the predictability to the DA setup are investigated. All of the sensitivity experiments performed in this thesis are started from the KENDA analysis based on IDEAL IC without large-scale uncertainty. Therefore, we drop the IC extension in the experiment names for increased simplicity in nomenclature.

Sensitivities to the Assigned Observation Errors and Type

First, we address the role of the assigned observation error in DA. In IDEAL-DAobserr3 (IDEAL-DAobserr05), the assigned radar reflectivity and velocity observation errors are tripled to 15 dBZ and 3 m s^{-1} (halved to 2.5 dBZ and 0.5 m s^{-1}). This means that observations get a significantly lower (higher) weight in the DA system, which, in case of IDEAL-DAobserr3, might be closer to a real system that faces issues of observation representativity and correlated observation errors that are often accounted

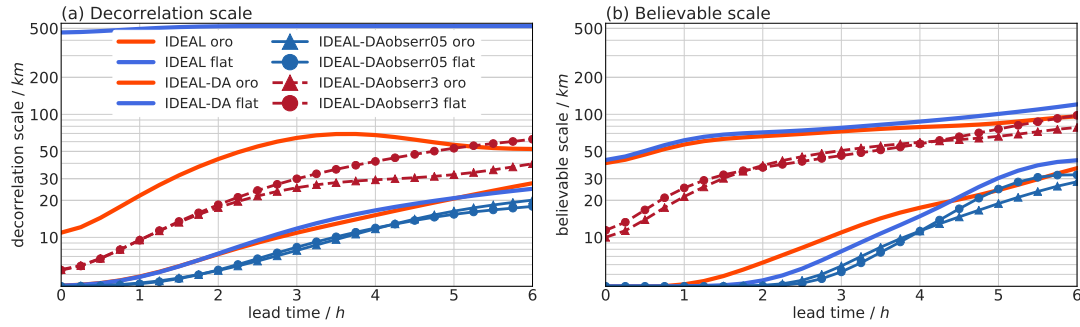


Figure 3.17: Time series of the decorrelation scale for IDEAL, IDEAL-DA and the sensitivity experiments IDEAL-DAobserr05 and IDEAL-DAobserr3.

for by inflating the assigned observation error (e.g., Bormann et al. (2003), Bauer et al. (2010)). IDEAL-DAobserr05 represents a further improved system and allows us to estimate its impact in the future. Improvement is not only possible regarding the density and quality of observations, but also in the DA. The assumption of uncorrelated observation errors, for example, is an active, recent research field (Weston et al., 2014; Bormann et al., 2016; Campbell et al., 2017), and its drop would result in adapted observation errors.

Already at the start of the forecasts, the simulations of IDEAL-DAobserr3 show less skill (cf. believable scale about 10 km in Fig. 3.17). The reduced weight of observations results in an analysis with more spread and larger errors. In contrast to IDEAL-DA, both believable and decorrelation scale begin to increase immediately after initialization. This results in larger predictable scales (meaning a larger proportion of small scales is unpredictable) for IDEAL-DAobserr3. Seen from the reversed perspective, a reduction of the errors assigned to the radar data by a factor of 3 (IDEAL-DAobserr3 versus IDEAL-DA) results in an increase in the forecast horizon by up to 3 h in the present study. However, it is important to note that the forecast horizon of IDEAL-DAobserr3 with large observation errors is still 3 h longer than in the IDEAL experiment without any DA. The believable scale in IDEAL-DAobserr3 amounts to 40 km at 3 h lead time and roughly 80 km at 6 h lead time, while in IDEAL it exceeds 40 km within the first hour and 80 km after 3 h. A further reduction in the observation error in IDEAL-DAobserr05 results in a further improved forecast, but the gain in forecast horizon is about 1 h smaller than before. Hohenegger and Schär (2007b) described similar rapid growth and propagation of different IC perturbations due to convective instability in real-case high-resolution forecasts, which is a sign of the inherently limited predictability of the atmosphere. In any case, our experiments are not designed to investigate those.

Secondly, only velocity or only reflectivity observations are assimilated in the experiments IDEAL-DAwind and IDEAL-DAreflectivity to investigate the relative importance of these observation types. Overall, the skill of the IDEAL-DAwind and IDEAL-DAreflectivity experiments is very similar. Both IDEAL-DAwind and IDEAL-DAreflectivity exhibit skill close to the assimilation of both observation types in IDEAL-DA. Both, the believable and decorrelation scale of IDEAL-DA, are never more than 20 km smaller than IDEAL-DAwind or IDEAL-DAreflectivity, which highlights that the

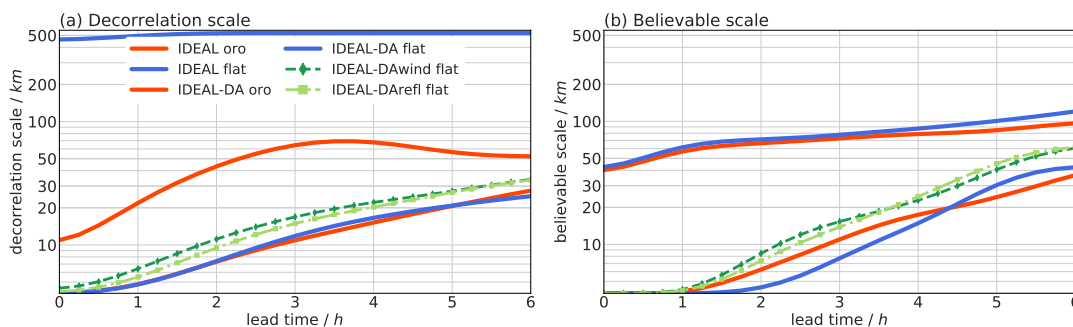


Figure 3.18: Time series of the decorrelation scale for IDEAL, IDEAL-DA and the sensitivity experiments IDEAL-DAwind and IDEAL-Dareflectivity.

assimilation of one observation type can already improve the forecast significantly (see Fig. 3.18). The beneficial impact on the results of only one assimilated observation type is promising, especially as it is already sufficient to compensate for the difference between *flat* and *oro* at short forecast ranges (not shown). Notably, the assimilation of wind velocities is technically more straightforward, as the required operators are much less complex as for reflectivity data (for example efficient modular volume-scanning radar forward operator (EMVORADO), Zeng et al., 2016).

In summary, the skill of all sensitivity experiments, apart from IDEAL-DAobserr05, lies in between the bounds set by the IDEAL and IDEAL-DA experiments. As IDEAL-DA represents a benchmark data assimilation experiment and in IDEAL no DA is applied, those experiments are expected to represent a lower and upper bound for the performed sensitivity experiments. The sensitivity experiments reveal that IDEAL-DAwind and IDEAL-Dareflectivity outperform IDEAL-DAobserr3 distinctly in both spatial metrics, hinting towards the importance of observation quality compared to the amount. IDEAL-DAobserr05 highlights that further improvements on the already excellent performance of IDEAL-DA are possible and the intrinsic limit of predictability is not reached.

Sensitivity to the DA Period

In the following paragraphs, the impact of radar DA at a later time within the life cycle of convection (experiment IDEAL-DAtime) is discussed. Since a shift in the DA window is equivalent to a time shift of the atmospheric conditions (Zhang et al., 2015), differences are to be expected. Compared to the benchmark experiments IDEAL-DA, the DA window in experiments IDEAL-DAtime is shifted by 4 h to 1100 – 1200 UTC when convection is well-developed and heaviest precipitation rates are observed (see Fig. 3.1c,g and Fig. 3.2).

The decorrelation and believable scale (see Fig. 3.19) of IDEAL-DAtime are only displayed for the first 4 h of the forecast (corresponding to 1200 – 1600 UTC), as IDEAL-DAtime suffers from sampling errors due to too few remaining thunderstorms in the last 2 h of the forecast (later than 1600 UTC). The believable scale and the decorrelation scale at the initial time are as good as for the IDEAL-DA experiments, but the initial

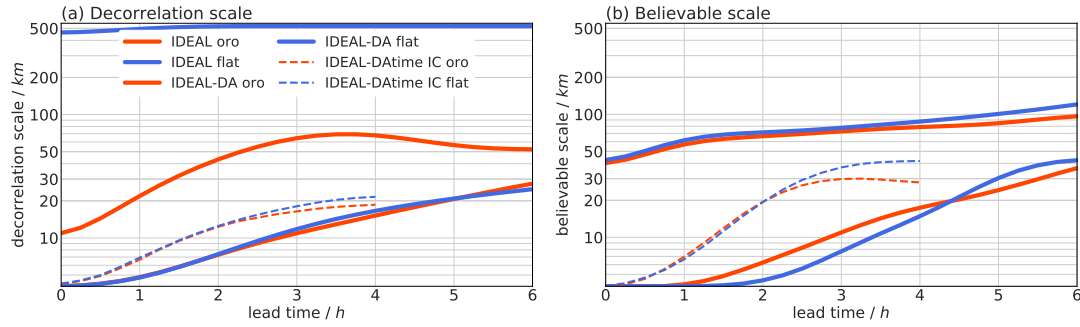


Figure 3.19: Time series of the decorrelation scale for IDEAL, IDEAL-DA and the sensitivities experiment IDEAL-DAtime.

errors grow more rapidly and already deviate from IDEAL-DA after 30 min. At 3 h lead time, the believable scale increases to 30 km, which is more than three times the believable scale of IDEAL-DA. The decorrelation scale amounts to 20 km at 4 h lead time corresponding to an increase of 30 % compared to IDEAL-DA. Both measures indicate a reduction of the forecast horizon by about 2 h when choosing the late data assimilation window.

It is evident that the stage of convective development, as well as the degree of the convective organization (further discussed in Section 3.3), play an essential role in the predictability limit. The convection-permitting COSMO model is able to predict the life cycle of convection with high accuracy from the moment convection is assimilated and thereby located correctly. If convection is assimilated in an early stage of development, it can, in a perfect model, potentially be predicted for its remaining lifetime of several hours. In the IDEAL-DAtime experiments, the convection is already in a later stage of development and its remaining lifetime is therefore shorter. As a result, predictability at a particular scale is shorter in IDEAL-DAtime than in IDEAL-DA. This finding is in agreement with earlier studies (Aksoy et al., 2009, 2010) showing that an EnKF performs differently for cases spanning super-cellular, linear, and multi-cellular organization of convection. A higher level of organization and coherency of the cells, resulting in a longer lifetime, likely leads to a longer-lasting impact of radar DA.

3.3. Impact of the Convective Mode in Idealized Experiments

It is an open meteorological question if the predictability of convection depends on its level of organization. If convection organizes itself in bands, multi-cellular systems or even a supercell, its predictability might be increased compared to smaller and unorganized convection because the higher level of organization leads to higher resilience against perturbations on the initial and boundary conditions. As a result, the convection lasts for an extended period and is less susceptible to the steering of the large-scale environment. Of course, this argument becomes only relevant if the NWP system predicts the right mode of convection at roughly the right location and point in time. Lilly (1990) speculated that large and organized convective systems are often relatively isolated and, therefore, only interact with a relatively calm environment. On the other hand, unorganized convection interacts with each other, continually exchanging mass and momentum, rendering prediction more challenging.

Aksoy et al. (2009, 2010) investigated the performance of an EnKF for cases spanning a convective band, multi-cellular organization, and a supercell. They find on the one hand that the DA produces analysis states of comparable quality in all cases and on the other side that the forecast ensembles started from those analyses diverge at different rates. As those studies focus on the performance of the EnKF, only lead times of up to 30 min are computed. Those short lead times might be influenced by the effects of model spin-up and do not allow for a thorough investigation into the predictability of those cases. Additionally, the performed forecasts are not comparable to the observations regarding precipitation amount and the lifetime of convection, which the author attributes to model deficiencies, especially microphysics in the supercell case, and the atmospheric soundings chosen to initialize the experiments. The results presented in the previous Section 3.2.2, regarding the decreased predictability in IDEAL-DAtime also motivate a further investigation, as different phases in a convective life cycle and different convective modes might look very similar.

The idealized OSSE setup introduced in Chapter 2 is uniquely suited to investigate the questions posed for several reasons. First, the radar DA ensures that mode, position, and timing of the convection in the ensemble forecasts are in line with the observations, which enables us to focus on the development of different convective modes. Secondly, the perfect model assumption of an OSSE eliminates the influence of model errors, which probably depend on the convective mode. Thirdly, the experiments are initialized from the same atmospheric sounding with different initial perturbations (see Fig. 2.2 for details), rendering the background conditions very similar. This is in contrast to previous case studies (Aksoy et al., 2009, 2010), which are based on soundings of different days with convection.

Figure 3.20 illustrates the setup for the experiments in this section in more detail. The Figure shows horizontal cross-sections of the maximum reflectivity within the model

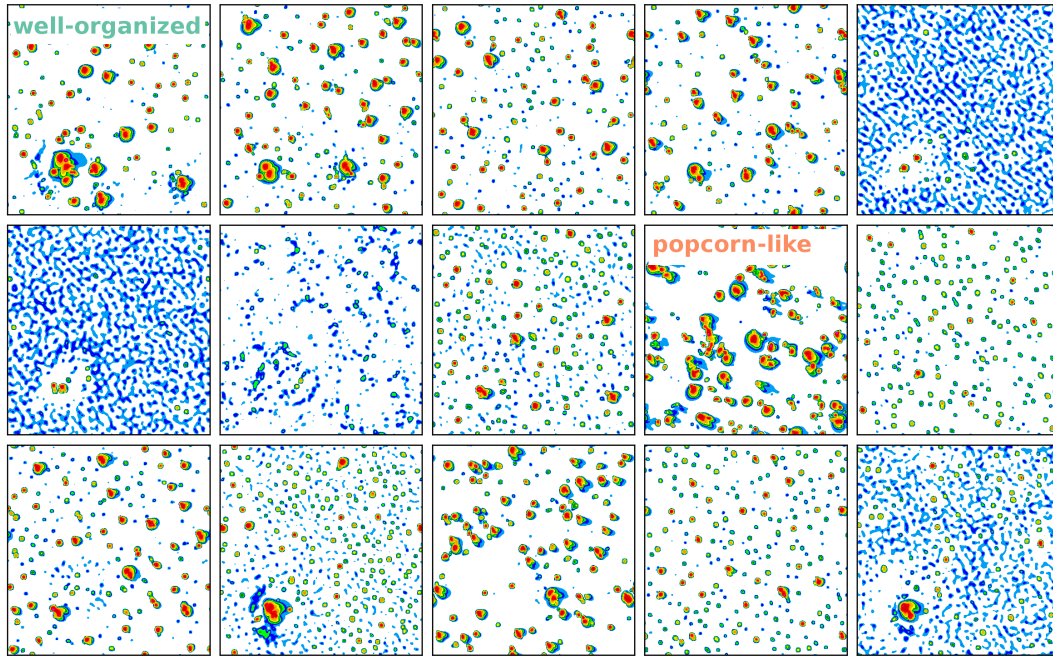


Figure 3.20: Horizontal cross-sections of the maximum reflectivity in the column of a sub-sample of ensemble members of IDEAL IBC *oro* valid at 0830 UTC or 30 min lead time.

column for a sub-sample of the 40 ensemble members of the IDEAL IBC *oro* experiment. Two of the shown ensemble members exhibit visually prominent and unique convective realizations and were chosen as Nature Runs providing *synthetic* observations for the DA cycling. Technically speaking, IDEAL-Nat and one ensemble member are exchanged before the DA cycling (see Fig. 2.3). This is possible as IDEAL-Nat shares the same configuration as the ensemble members apart from the noise seed used for initialization. The resulting DA ensembles are forced in the convective modes of their respective Nature Runs, whose evolution after that can be seen in Fig. 3.21. Apart from the exchanged Nature Run, which means the assimilated observations, the setups of COSMO and KENDA are identical to the IDEAL-DA IBC *oro* experiment described in the previous section. We refer to these ensembles as IDEAL-DA *well-organized* and IDEAL-DA *popcorn-like*. As both experiments feature orography and large-scale uncertainty, we drop the *oro* and IBC for convenience and clarity of the terminology. Otherwise, the naming convention is in line with previous chapters and summarized in Table 4.3.

IDEAL-Nat *popcorn-like* exhibits small and medium-sized convection that is evenly distributed within the domain (cf. Fig. 3.21). Those cells do not organize to larger systems throughout the experiment. In contrast, IDEAL-Nat *well-organized* depicts strong convection triggered in the lee of the orography, which organizes itself quickly to a MCS and suppresses further convection in parts of the domain. In this simulation, only the large MCS and a few strong cells persist throughout the simulation.

The Spaghetti plots of IDEAL-DA *well-organized* and IDEAL-DA *popcorn-like* in Fig. 3.22 highlight the positive impact of the radar DA and the efficiency of the cycling in shifting the convection in all ensemble members in the organizational mode of the

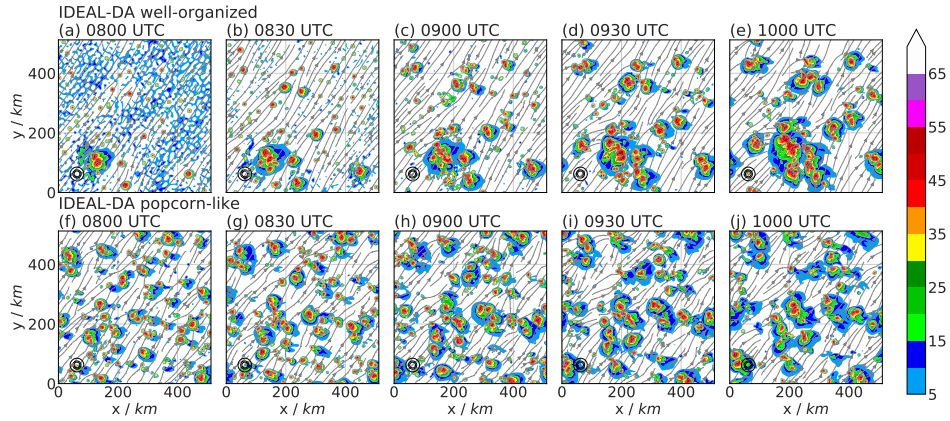


Figure 3.21: Maximum reflectivity in the column of the two investigated Nature Runs *well-organized* (a–e) and *popcorn-like* (f–j); the grey streamlines depict the horizontal wind at 1500 m above ground; plots are valid at 0800 UTC (a),(f), 0830 UTC (b),(g), 0900 UTC (c),(h), 9300 UTC (d),(i) and 1000 UTC (e),(j); the black circles represent the location and height of the orography.

respective Nature Run. Especially at the beginning of the forecasts at 0800 UTC, most convective cells in the respective Nature Runs are captured by the ensembles and the spatial spread seems visually comparable, although IDEAL-DA *popcorn-like* shows more false positives. For the first 2 h lead time, the convection in the DA ensembles diverges slowly, but IDEAL-DA *well-organized* seems to agree with its Nature Run to a higher degree than IDEAL-DA *popcorn-like*.

So far, the level of convective organization in the two experiments was only assessed visually. To quantify our impression, we counted the number of convective cells with a simple counting algorithm that identifies coherent structures above the thresholds of 19 and 45 dBZ within the domain. The number of convective cells is not a direct measure of the level of organization, but it is an easily computed indicator and directly connected to the hypothesis of Lilly (1990). In a statistical sense, more organized convection tends to reach higher parts of the troposphere, be more long-lived, and produce higher values of reflectivity and precipitation. Therefore, a higher number of strong cells and less variability of the count hints towards more organized convection.

Figure 3.23a shows a very similar number of predominantly weaker cells above the lower threshold of 19 dBZ (approximately 1 mm/h precipitation). Especially within the first 1 h lead time, over 100 cells are counted. The large number quickly adjusts as those small cells merge. In contrast, Figure 3.23b reveals a striking difference between the experiments in the number of strong convective cells. The number of strong cells in IDEAL-DA *well-organized* grows from little over 20 to around 60 within 90 min and decreases slowly throughout the experiment with around 30 cells at 6 h lead time. IDEAL-DA *popcorn-like* exhibits a higher variability in the number of strong cells, which peaks around 90 cells within the first hour and drops steadily to just 10 cells.

Another way to illustrate the variability of the convection in the experiments are spectra. Therefore, we computed the 2-dimensional spectral density of the reflectivity fields of

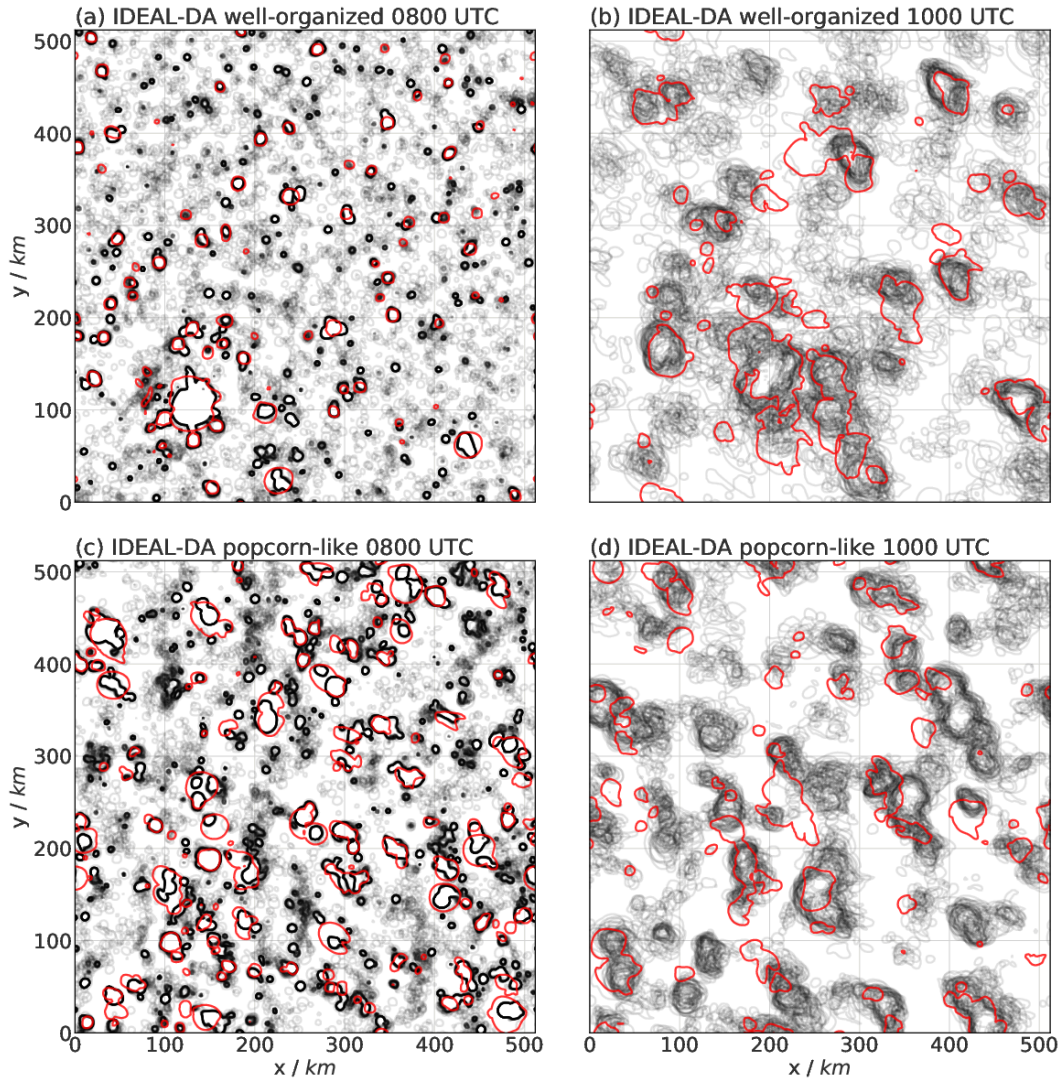


Figure 3.22: Spaghetti plots of IDEAL-DA *well-organized* (a,b) and IDEAL-DA *popcorn-like* (c,d) at 0800 UTC / 0h lead time (a,c) and 1000 UTC / 2h lead time (b,d): isolines of 19 dBZ of all 40 ensemble members (black) and the respective deterministic Nature Run (red).

every ensemble member with the FFT. Figure 3.24 shows those spectra, averaged over the ensemble dimension, for different lead times. The general features of both experiments are very comparable, with the largest values at larger scales of around 100 km. The striking difference is the continuous decrease of the spectral density of about an order of magnitude in IDEAL-DA *popcorn-like* on all scales below 100 km, which is much less pronounced in IDEAL-DA *well-organized*. Although the average spectral density over time of the experiments is comparable the variability is more pronounced in IDEAL-DA *popcorn-like*, showing that the convective cells are more organized, structured, and therefore long-lived in IDEAL-DA *well-organized*.

The decorrelation scale in Fig. 3.25a shows comparable predictability of the model state in the two experiments (IDEAL-DA *well-organized* and IDEAL-DA *popcorn-like*),

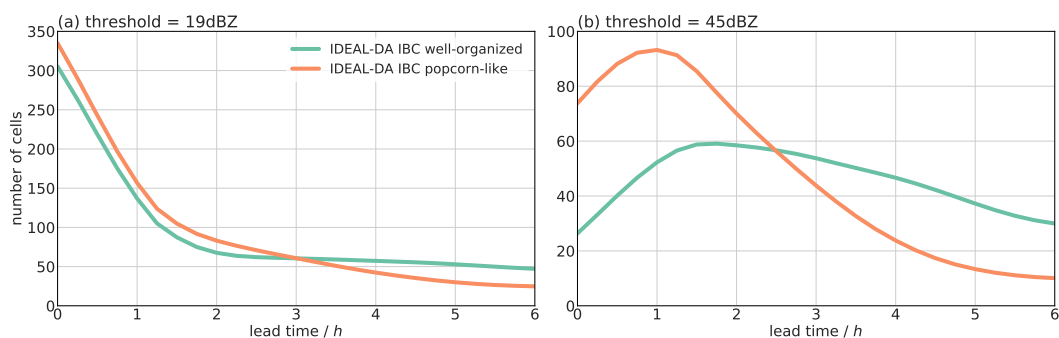


Figure 3.23: Time series of the number of cells present in the domain for thresholds of 19 and 45 dBZ. A cell is a consecutive area of grid boxes above the respective threshold.

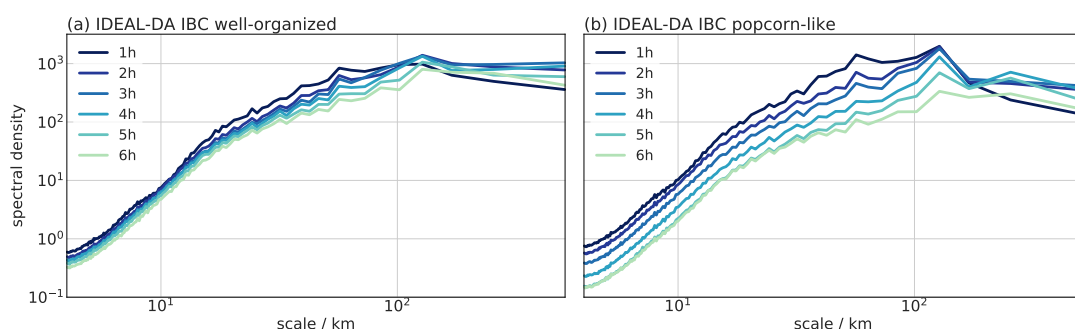


Figure 3.24: Spectra of cross-sections of the maximum reflectivity fields of IDEAL-DA *well-organized* (a) and IDEAL-DA *popcorn-like* (b). The coloring indicates the lead time.

growing from under 10 km at 0 h lead time to about 35 km after 6 h, which is also almost identical to IDEAL-DA IBC *oro*. The decorrelation scale indicates that the analysis states are comparably good and that the NWP model produces a comparable spread in the location of the convective cells throughout the simulations.

In contrast, the believable scale (cf. Fig. 3.25b) depicts a rapid loss of predictable scales within the first 60 min lead time in IDEAL-DA *popcorn-like*, raising (i.e. worsening) from 25 to over 60 km. After the initial jump, the predictable skill difference between IDEAL-DA *well-organized* and IDEAL-DA *popcorn-like* grows steadily to up to 80 km after 6 h lead time.

The different behavior of the two metrics is rooted in the fact that only the believable scale includes a comparison to a Nature Run (cf. Sec. 2.5.4) and has some interesting implications. The Nature Run is set up in the same way as the ensemble members, apart from the initial noise seed. The only difference is that no DA was applied to it, rather did itself source the observations for the ensemble. The believable scale highlights a rapid divergence of the whole IDEAL-DA *popcorn-like* ensemble from the Nature Run, while the decorrelation scale signals a similar ensemble divergence. The difference we see is not a different predictability of the model state, but a different ‘model predictability of the analysis state’ and it is comparable to the results of Aksoy et al. (2009, 2010). Systematic model deficiencies in weather forecasts can lead to similar situations in

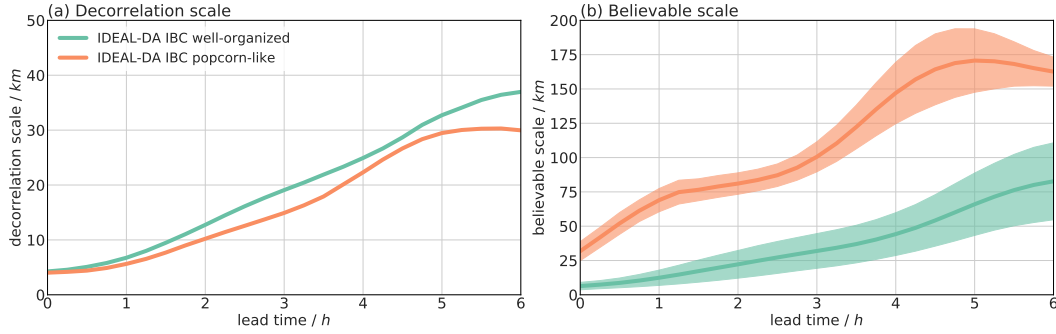


Figure 3.25: Time series of the decorrelation and the believable scale of the IDEAL-DA *well-organized* and IDEAL-DA *popcorn-like* experiments.

which the truth lies outside the ensemble space. IDEAL-DA *popcorn-like* represents such a case in an idealized OSSE.

The, in a dynamical sense, more balanced analysis of IDEAL-DA *well-organized* leads to reduced spurious convection in the forecast (Lange et al., 2017) and is in our opinion the main contributor to the increased predictability of IDEAL-DA *well-organized*. Several other factors could in general be important, but would also influence the decorrelation scale. From a meteorological standpoint, the lifetime of organized convection is enhanced. In an idealized setup with radar DA fixing the initial position and structure of the convection, this could extend the forecast horizon, as we showed in the DAtime experiment in Sec. 3.2.2. Furthermore, from an error growth perspective, the structure and coherence of organized convection could limit the growth of small initial errors and lead to a reduced susceptibility towards those unavoidable IC errors. Finally, in IDEAL-DA *well-organized*, fewer nonlinear interactions between thunderstorms have to be represented, thereby potentially improving the model forecast.

In contrast to previous studies, we extended the scope of this study towards the evolution and predictability of different modes of convection. The believable scale revealed increased practical predictability in the presence of organized convection (IDEAL-DA *well-organized*) compared to an unstructured situation with more randomly occurring convection (IDEAL-DA *popcorn-like*), where the level of organization was determined by the number of cells in the domain and the variability of the spectral density.

Our experiments confirm the findings of previous publications regarding the performance of the LETKF in different convective situations. The Spaghetti plots and the initial values of decorrelation and believable scale show that the convection in the analysis state for both very different situations shifts towards—an encouraging result. But we also observe problems of the forecast model to reproduce the Nature Run IDEAL-Nat *popcorn-like* for an extended period, which could be related to an unbalanced analysis state. As from a forecaster’s perspective, a functional analysis is an analysis that produces a good forecast; this is an issue to be addressed in the future.

Unfortunately, an estimation of the relative importance of the abovementioned factors is out of the scope of this work. Nevertheless, the results presented in this section show, to our knowledge for the first time, the influence of the organization level of convection on its predictability in a comprehensive way.

3.4. Predictability of Convection in Operational NWP

In this section, we move from the idealized OSSE setup to different configurations of the operational COSMO system at DWD, which also implies that the perfect model assumption is not valid anymore. This includes, for example, systematic model errors, errors in the forward operator used in the DA scheme, as well as issues with measurement errors and observation representativity.

We can estimate the combined impact of these effects by comparing the decorrelation and believable scale. The predictability of the model state is provided by the decorrelation scale, which is an intra-ensemble measure and, therefore, insensitive to systematic differences to the atmosphere. In contrast, the believable scale picks up those differences and penalizes them. It thus provides an estimate of the model predictability of the atmospheric state. In summary, the decorrelation scale is a proxy for the intrinsic predictability, and the comparison to the believable scale estimates the potentially possible forecast improvement if the NWP system is improved to perfection.

The configurations chosen allow us to reintroduce inherent complexity in the simulations in two steps. First, we analyze a HIW period in Germany that features a very similar weather situation as the idealized experiments, we presented so far, but features with terrain over Germany. In a second step, we look into operational forecasts for the summer periods 2014–2016 in the same domain. The more extended period guarantees that a wide variety of weather situations is included in the data, thus allowing for an additional inquiry in the dependence of convective predictability on orography, the weather regime and the interplay of both.

The range of NWP systems in this section, consisting of COSMO coupled with ICs based on a multi-model ensemble with physics perturbations, KENDA with LHN and KENDA with 3-dimensional direct data assimilation, also gives us an impression of the predictability limits of the old, the current and the future operational DWD systems.

3.4.1. COSMO KENDA: Impact of Orography and NWP system

First, we apply our scale-dependent predictability measures to weather forecasts of a convectively active, HIW two weeks period between 27 May and 10 June 2016 in Germany (Piper et al., 2016). In contrast to the idealized experiments in the previous chapter, the metrics are applied to the hourly aggregated precipitation fields. The time aggregation will impact the scores slightly positively, as it accounts for timing

errors in the precipitation. Apart from that, scores for hourly precipitation instead of instantaneous column-maximum reflectivity showed comparable results in previous studies (Stratman et al., 2013). We compare two approaches, namely LHN and direct 3-dimensional assimilation of the radar reflectivity and radar radial wind. Note that the wind information is disregarded in the LHN. Additionally, we investigate the influence of the mixing length, which is related to convective triggering. We vary the mixing length between the operational value of 150 m and a value of 500 m, which is considered to be more realistic.

Impact of Orography

Similar to Seifert et al. (2012) and Kühnlein et al. (2014), we divide Germany in a Northern and Southern part to represent regions that are orographically influenced to a higher or lower degree (see Fig. 2.5). For the remainder of this section, results will be shown for those parts to assess orographic influence over Germany. First, we focus on the DE-DA-LHN experiment and Fig. 3.26a,c, showing results for a mixing length of 500 m before we address the DE-DA-3DRad experiments and the mixing length of 150 m.

Figure 3.26a depicts the development of the decorrelation scale with lead time. As the forecasts are initialized hourly every day of the HIW period between 1000 and 1800 UTC and run for 6 h, the figure depicts averages over lead time that are valid for different times of the day. This is necessary as the effect of the lead time dominates over the diurnal cycle. The black line indicates the development over the entire domain, while blue and red represent the North and South Half. The decorrelation scale grows continuously from below 10 km at 1 h lead time to 40 km in the South Half (almost 70 km in the North), while the values of the whole domain remain within those bounds, but closer to the South. In our study, the orographically more influenced South shows significantly higher skill throughout the forecasts and the difference between North and South is increasing with lead time.

As in the idealized setup, we interpret these observations as a reliable indicator for increased predictability of the model state of deep convection in Southern Germany due to orography. Convective precipitation over orographic regions is, in a probabilistic sense, more constrained to those regions favored by the interaction of orography and, for example, synoptic-scale flow or radiation, increasing the predictability in these regions. It is, however, striking that the findings of the idealized setup can be reproduced in the presence of model errors.

As the decorrelation scale only assesses inter-ensemble variability and involves no comparison to observations, the believable scale offers additional information about the practical predictability, more specifically the model predictability of the atmospheric state, in the COSMO-KENDA experiments (see Fig 3.26c). It is computed for the 75 percentile, a variable threshold to account for differences in the precipitation amount in the different regions and over the period.

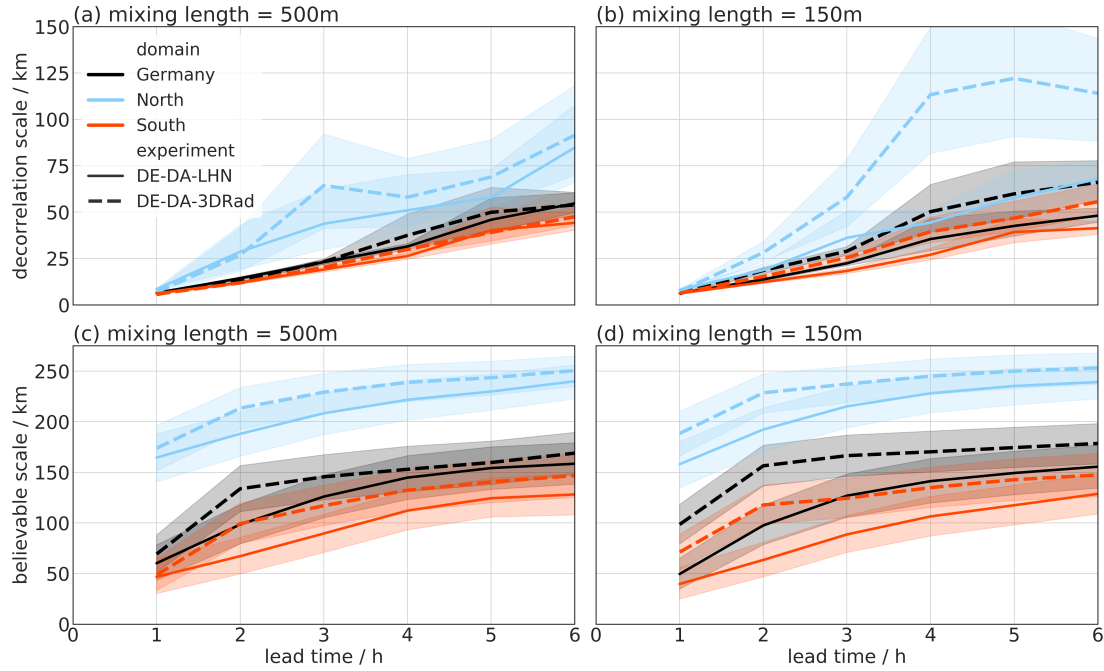


Figure 3.26: Time series of the decorrelation scale (a,c) and the believable scale for the 75th percentile as threshold (b,d) for the DE-DA experiments (between 27 May and 10 June 2016) with mixing length 150 m (a,b) and 500 m (c,d). The shading represents the 95% confidence intervals determined by bootstrapping ($n=10000$).

The believable scale depicts the same trend as the decorrelation scale with continuous growth. However, the growth starts to saturate within the lead time of 6 h. As expected for a comparison to observations, the believable scale is larger (by a factor of 2–3) than the decorrelation scale, which is in line with previous studies describing convection-permitting models as overconfident and under-dispersive (Romine et al., 2014; Dey et al., 2016). The difference between the orographic South and the comparatively flat North appears even more significant and amounts to 100 km after 6 h lead time.

The fact that the mean of the values of North and South Germany are unequal to the believable scale of entire Germany can be attributed to the non-linearity of the FSS, which means $FSS(Ger) \neq (FSS(N) + FSS(S))/2$. A larger precipitation object, for example, imposes a more significant and disproportional influence on the FSS (Skok and Roberts, 2018).

The believable scale highlights that the investigated short-range forecasts show higher skill and model predictability of the atmospheric state over Southern Germany related to the presence of orography. It also shows that the skill of state-of-the-art convective-scale ensemble forecasts converges towards a climatological value, defined by the weather regime and region, within only 4 to 6 h lead time.

The experiments in the HIW period allowed us to extend our findings from the idealized setup (see Sec. 3.1) to the pre-operational and operational LAMs of DWD. We find in both realistic configurations (DE-DA-LHN and DE-DA-3DVar) with complex terrain and a weather situation comparable to the idealized simulations that precipitation is

more predictable in the orographic South than in the comparatively plain North of Germany. This can be seen in increased predictability of the model state measured by the decorrelation scale, but also in increased model predictability of the atmospheric state, measured by the believable scale.

A comparison between the idealized and the DE-DA forecasts reveals another feature. In the idealized experiments, DA was able to compensate for the positive impact of orography on the predictability for several hours by providing excellent IC and correcting flawed LBC. In the more realistic and complex simulations over Germany, the scores of North and South do not collapse after DA, highlighting the further potential to improve the assimilation of radar observations and subsequently precipitation forecasts, especially in Northern Germany, where less orography and therefore fewer triggers of convection are present. It also hints to the fact that in the operational system, orography does not just represent a trigger for convection, but also a source of model errors related to the complex lower boundary.

Impact of Radar DA and Mixing Length

For a mixing length of 500 m, the DE-DA-3DVar experiments perform similarly or only slightly less well than DE-DA-LHN in all domains measured by decorrelation and believable scale. This is an encouraging result, as the LHN has been used operationally at DWD for almost a decade, while the 3D radar data assimilation is a new system, which includes features in development like the warm bubble triggering (Zeng et al., 2019). Additionally, the 3D radar DA is only applied every 60 min for the radar scan valid at this time, while the LHN, as a part of the COSMO model at DWD, is invoked in every time step in the model (with radar data interpolated between the corresponding radar scans performed in five-minute intervals), which has a potentially beneficial impact. To compare the performance of the systems, we also have to consider that the radial wind information measured by the radar is disregarded in the LHN, but assimilated in DE-DA-3DVar.

The results for the decorrelation scale of different mixing lengths (500 and 150 m) are displayed in Fig. 3.26a,b. We can see a higher variability between the experiments and the different domains, especially in the North. We find the remarkable result that direct 3D radar DA is impaired by the smaller mixing length for lead times larger than 2 h, reflected in a larger decorrelation scale especially for the less orographically influenced North of Germany. In contrast, we find that the DE-DA-LHN with mixing length 500 m performs slightly less well with the more realistic mixing length, probably due to its tuning to the mixing length of 150 m that is used in the operational system. The believable scale in Fig 3.26d is unable to pick up this feature as the comparison to observations already draws the predictable scale towards higher 'climatological' values representative for region and weather situation.

The decorrelation scale of DE-DA-3DVar can explain the effect of the mixing length on the model in the North in combination with a mixing length of 150 m. A smaller mixing length in the boundary layer scheme causes fewer mixing, leading to stronger

temperature gradients and earlier triggering of convection. In Southern Germany, the orography provides triggers for initializing convection in areas with favorable conditions. In contrast, in the North, the absence of such triggers results in substantial spatial uncertainty of convection.

The two sets of DE-DA experiments (with LHN and 3D radar DA) exhibit overall good and comparable results, which is encouraging as LHN has been operational for almost ten years and the 3D radar DA includes new and experimental features. However, the sensitivities of different DA schemes to the mixing length, a parameter in a physical parametrization of the forecast model, are substantial and call for a careful and combined development of forecast model and data assimilation ((Gustafsson et al., 2017)).

3.4.2. COSMO-DE-EPS: Impact of Orography and Weather Regime

As a final generalization, we reintroduce the natural variability of the weather situation after we had already included the complex orography in the last section. We aim to substantiate our findings in a more statistical sense by investigating three consecutive summer periods (May–September) in the then operational DE-EPS system. For the examination, we rely on 27 h forecasts, started at 0000 UTC and 1200 UTC every day. The initial and boundary conditions in the 20 members ensemble are created by a combination of down-scaling from multiple models and perturbations in parametrization parameters (see Sec. 2.4.2). The focus of the following experiments is the evolution of convective predictability under different synoptic conditions and in regions that are more or less orographically influenced, as well as the interplay of these effects.

Figure 3.27 depicts the time development of the NRMSE averaged over the complete period. As forecasts are started at 0000 UTC, the lead time and the time of the day are identical. In contrast to the previous experiments, the amount of available data, around 1000 forecasts after quality control, allowed us to split the results into two different weather regimes defined by the convective timescale. We found that about 30 % of the cases are locally forced.

The forecasts up-scaled to 8 km, which is smaller than the effective resolution of the model, but identical to the up-scaling in the idealized simulations, show in general poor forecast skill. The entire domain and the sub-domains are indistinguishable, while the synoptic and local forcing situations depict distinctly different scores. Measured by NRMSE, the synoptic forcing regime is practically more predictable than the locally forced one for an up-scaling of 8 km. The same holds for the two up-scaling values of 32 km and 128 km.

Additionally, the score gradually improves as larger areas are verified. Similarly, the rate at which the NRMSE deteriorates becomes smaller for larger verification areas. This can probably be attributed to the fact that atmospheric processes on scales of

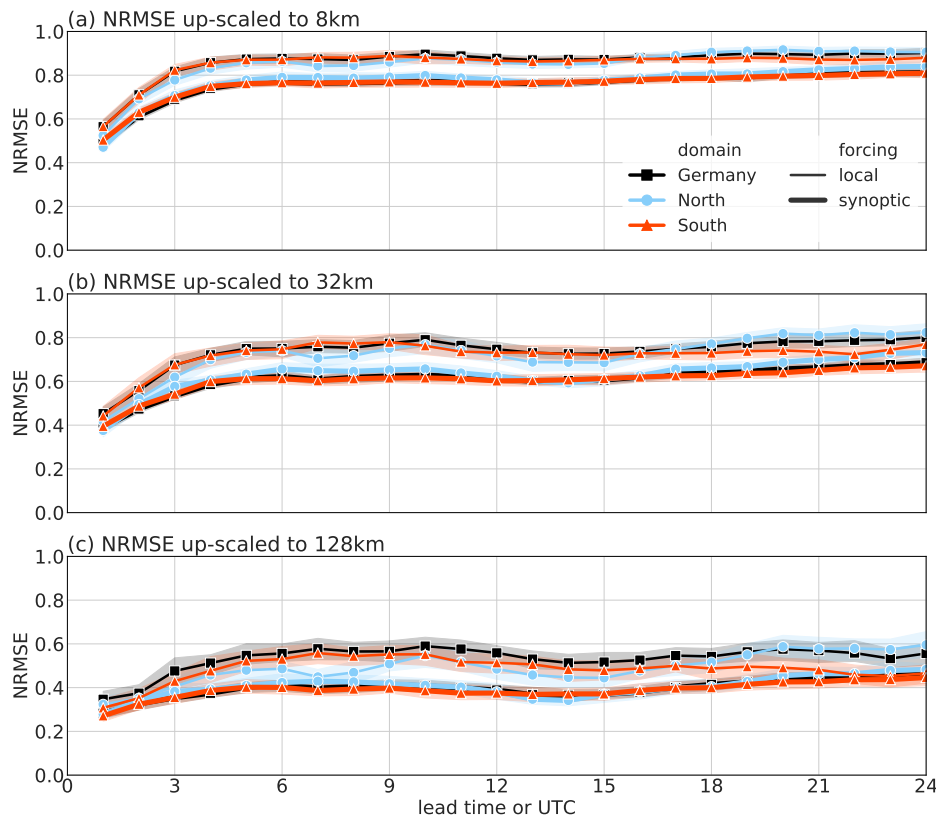


Figure 3.27: NRMSE of the hourly accumulated precipitation for three levels of up-scaling (8, 32, 128 km); the results are shown for the complete domain of Germany (black), and split in the Northern (red) and Southern (blue) sub-domains. Additionally, they are split by the convective adjustment timescale in synoptically and locally forced situations. The shading represents the 95 % confidence intervals determined by bootstrapping ($n=10000$).

128 km or above are in general slower than on 32 km. The forecast horizon based on the NRMSE is also dependent on the upscaling and ranges in the order of a few hours.

In Figure 3.28, the decorrelation and believable scale for the experiments already discussed employing the NRMSE (see Fig. 3.27) are shown for initialization times at 0000 UTC, and additionally, at 1200 UTC. Figure 3.28a depicts the time development of the decorrelation scale for the forecasts started at 0000 UTC averaged over the complete period (the lead time and the time of the day are identical in this case).

Within the first 12 h lead time, before local noon, the decorrelation scale remains below 25 km for both synoptic regimes. After that, around 1200 UTC, the increase of the decorrelation scale accelerates in both regimes. We attribute the rapid growth of forecast errors between 1200 and 1800 UTC to the higher frequency of deep convection during the day, which is a significant contributor to error growth in the atmosphere

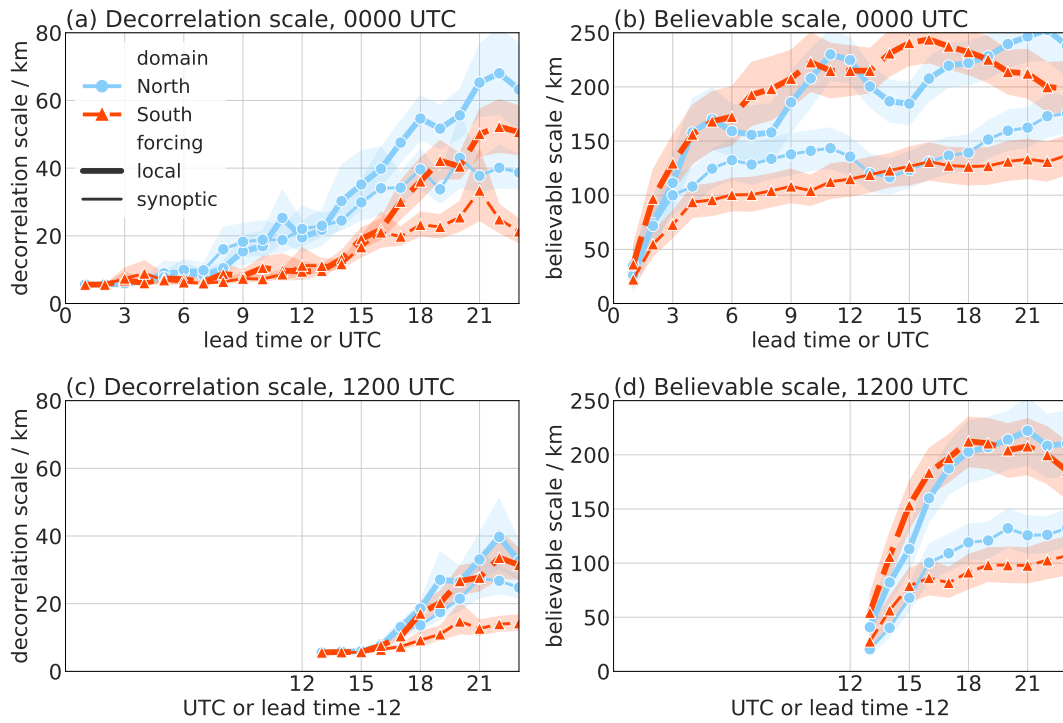


Figure 3.28: Averaged time series of the decorrelation scale (a,c) and the believable scale for the 75th percentile as threshold (b,d) for the COSMO-DE-EPS experiments issued at 0000 UTC (a,b) and 1200 UTC (c,d) in summer 2014–2016. As forecasts are started at 0000 UTC the lead time and the time of the day are identical for (a,b) and shifted by 12 h in (c,d). The shading represents the 95 % confidence intervals determined by bootstrapping ($n=10000$).

(Zhang et al., 2007; Selz and Craig, 2015; Sun and Zhang, 2016). Especially in situations characterized by local forcing, the decorrelation scale grows rapidly and reaches around 60 km at the end of the forecast. Therefore, the unpredictable scales become two to three times as large if no synoptic forcing acts to organize the convection. Keil et al. (2014) and Kühnlein et al. (2014) also describe higher practical predictability of synoptically compared to locally forced situations.

Additionally, we separated the forecasts in both weather regimes into a North and South domain. In both investigated regimes, the predictability of convection in the North and South part of Germany (representing flat and orographic terrain) is identical at early lead times before it begins to diverge. In both synoptic situations, the South depicts a decorrelation scale that is roughly 20 km smaller, which means those scales remain predictable for a longer time than in the North. Again, we interpret the consistent difference between the North and the South as increased predictability of the model state in the South caused by the orography.

The decorrelation scale for the forecasts initialized at 1200 UTC is shown in Fig. 3.28c. In those forecasts, the decorrelation scale starts to grow later, around 1600 UTC, in terms of local time, but in terms of lead time, the predictability of the model state is lost earlier due to the active convection in the afternoon. As do the forecasts started

at 0000 UTC, the 1200 UTC forecasts depict higher predictability of the South in synoptically forced situations, while no difference is visible for local forcing conditions.

At this point, it is interesting to compare the decorrelation scale of DE-EPS, initialized with downscaled ICs from global models, under local forcing conditions to the IDEAL-DA IBC *oro* experiment. The local forcing conditions are more representative for the sounding we used to initialize the idealized simulations, and the decorrelation scale of IDEAL-DA IBC *flat* only depicts the domain size. The believable scale also cannot be compared, as in the idealized setup no model error is present that would cause a systematic difference between observation and model as in the COSMO-DE-EPS. The local forcing DE-EPS experiments, initialized at 0000 UTC, depict a growth rate of around 30 km in 6 h in the convectively active period between 1200 and 1800 UTC. The growth rate of the decorrelation scale for the 1200 UTC forecasts is smaller in every given six hours period, while the idealized experiment IDEAL-DA IBC depicts a growth rate around 50 km in 6 h lead time (see Fig. 3.16). The reduced growth rate of the decorrelation scale hints towards an under-dispersive ensemble, as described by Romine et al. (2014) and Dey et al. (2016).

The believable scale (see Fig. 3.28b) highlights features different from those of the decorrelation scale. A comparison of the forecasts to observations reveals a rapid loss of predictability of the atmospheric state within the first 3 h–5 h independent of the initialization time, a period in which the decorrelation scale still picked up a high agreement within the ensemble. The comparison to the decorrelation scale emphasizes a too small ensemble spread that is not representative of the uncertainty of the situation and does not capture the precipitation observed. In other words, the forecasts are, as already suspected, under-dispersive and overconfident. After the initial rapid growth period to around 120 km in the synoptic regime and 200 km in the local forcing regime, the believable scale reaches an upper limit that represents the climatological predictability of deep convection in summer in Central Europe. Although the believable scale behaves similar in terms of lead time for forecasts initialized at 0000 UTC and 1200 UTC, the forecasts initialized later still show higher skill than those initialized in the night hours.

In the investigated operational COSMO-DE-EPS forecasts, the predictability limits are dominated by the weather regime, while the orography plays no significant role in the local forcing regime and, possibly, a minor role with synoptic forcing (see lead times of 4 to 10 h in Fig. 3.28b,d). Baur et al. (2018) describe a similar finding concerning the impact of heterogeneous soil moisture perturbations, which had only a minor influence in synoptically forced situations. The above-mentioned upper limit for local forcing regimes is insensitive to the threshold of the FSS the believable scale is based on. For the synoptic regime, the choice of the threshold shifts the saturation level, but the characteristics of the plots remain unchanged.

Conclusions

With the availability of the necessary computing power, convection-permitting Ensemble Prediction Systems (EPSs) became operational at most operational weather centers within the last 15 years. Their advent constitutes a step-change in our ability to forecast convection (Clark et al., 2016). Due to their immediate consequence to our day to day lives, precipitation forecasts are of great interest to the general public. Of course, forecasting deep convection and thunderstorms is a fundamental task of operational weather centers, as they can be accompanied by heavy precipitation, wind gusts, and destructive hail. They also have a considerable socio-economic impact and can even pose a risk to human life.

The predictability of the atmosphere is inherently limited, especially on convective scales and in the presence of deep convection, rendering forecasts of convective precipitation a challenging task. However, so-called sources of predictability, such as orography or the prevailing weather regime, could provide the means to extend the forecast horizon on the convective scale. The identification of predictability limits of deep convection and their dependence on the abovementioned factors in operationally used state-of-the-art NWP systems is necessary to evaluate the performance of current models and build confidence in the forecasts. Those state-of-the-art systems rely on, or will rely in the near future on, the assimilation of radar data, which renders estimates of its potential for future developments critical at this time. Accordingly, it is of interest to extend our study and include radar DA, especially as its effect within the NWP system is similar to the influence of orography—it provides information on the location of convection.

The identification of predictability limits of deep convection, however, poses significant challenges. The complexity of the problem lies both within the intermittent, spatio-temporally highly variable nature of convective precipitation depending on an ever-changing weather situation, and within the different components of modern convective-scale NWP systems. Those include the challenge to assimilate high-resolution observations to provide the initial state of the atmosphere, as well as an appropriate scale-dependent verification.

We addressed those issues with a hierarchy of COSMO configurations of different complexities ranging from an idealized setup to operational DWD forecasts, which are coupled with the DA system KENDA or the latent heat nudging (LHN) that is operational at this point. In addition, a combination of sophisticated metrics addresses

specific facets of practical predictability, namely the predictability of the model state and the model predictability of the atmospheric state.

The idealized OSSE reduces the complexity and enables us to isolate the effects of orography and radar DA, as it allows simulations that are conceived to address very specific questions. We therefore simulate convection in the presence of simple orography and exclude the inherent weather situation uncertainty by prescribing horizontally homogeneous atmospheric conditions with periodic boundaries susceptible to long-lasting convection in Central Europe. At the same time, we neglect approximations and inconsistencies in the description of atmospheric processes in the NWP model by using a perfect model approach.

The operational and pre-operational forecasts then reintroduce realism in the form of complex orography in a domain over Germany. First, we investigate practical predictability limits in a two weeks HIW period that features a similar convective environment as the OSSE but complex terrain, and secondly extend the scope even further by analyzing three consecutive summers from 2014 to 2016. The long period includes a variety of weather situations and enables us to conclude in a statistical way.

While this thesis can by no means comprehensively address all open questions regarding the predictability of deep convection, its novelty lies in its seamless approach and in the application of scale-dependent verification metrics complementing each other. As a result, the present thesis relied heavily on systems and tools of convective-scale forecasting developed in recent years to answer the research questions presented in the introduction:

- How do sources of predictability, such as orography, influence the predictability of convective precipitation?
- What is the potential impact of radar data assimilation on the practical predictability of convective precipitation, and is its impact affected by orography?
- What is the impact of the convective mode or the level of organization on the predictability?
- In which way do sources of predictability, like orography or the synoptic weather regime, impact the practical predictability of deep convection in an operational NWP system over Germany these days?

We will summarize the findings of this work in the next two sections structured as follows. First, the effects of orography and the synoptic weather regime on the predictability of convective precipitation are shown based on the idealized, perfect model environment, a two weeks HIW period, and three years of then operational ensemble forecasts of Germany in the summers of 2014 to 2016.

Secondly, we address the effect of radar DA and its interplay with orography in an idealized OSSE and the operational and pre-operational COSMO-KENDA systems of

DWD in the same two weeks period. We also summarize sensitivities to observation errors, observation type and the time of assimilation in the idealized case, and compare the performance of LHN, direct 3-dimensional radar data assimilation and COSMO-DE-EPS. The effect of the convective mode is also discussed here, as it revealed a strong connection to the DA. Finally, we provide a brief outlook into future applications of our results and how they spark new questions that can be investigated in a similar framework.

4.1. Orography and Weather Regime

Nature provides sources of predictability for deep convection. The lower boundary condition and, particularly, orography represent one prominent source of predictability of convective precipitation. Anthes (1986) hypothesized that orographic features hand down predictability to the storm scale and increase their predictability. As a result, the probability of deep convection in the presence of orography becomes a function of the complex terrain and the flow (Berri and Paegle, 1990). This is corroborated by Laprise et al. (2000) stating that LBCs in combination with strong surface forcings, such as orography, are able to correct flawed ICs to some degree. However, Nuss and Miller (2001) and Picard and Mass (2017) report a sensitivity of orographically influenced local precipitation patterns to perturbations of the synoptic-scale flow direction. This study can confirm both findings—increased predictability of convection and a crucial sensitivity to the synoptic-scale uncertainties.

The significance is considerable, as a substantial portion of deep summer convection in mid-latitude regions of the world can be linked to orographic features. Carbone and Tuttle (2008) find that about 60% of midsummer rainfall between the Rocky Mountains and the Appalachians is caused by propagating rain systems triggered by elevated terrain. Levizzani et al. (2010) show similar results for the Mediterranean.

In the OSSE, we find that a single Gaussian mountain visually changes the distribution of convection within an ensemble and increases the predictability of convection measured by NRMSE, decorrelation and believable scale. The effect is more pronounced in the vicinity of the mountain and for longer lead times, as ICs lose relative importance, while the orography acts as a continuous forcing (Vié et al., 2011; Kühnlein et al., 2014). In contrast to previous studies (Duda and Gallus, 2013; Isotta et al., 2014; Kovacs and Kirshbaum, 2016; Gowan et al., 2018), our idealized setup without complex orography and horizontally homogeneous initial conditions facilitates a distinct attribution and quantification of orographic effects on the spatial distribution of precipitation.

However, realistic synoptic-scale uncertainties, which are unavoidable in a Limited-Area Model (LAM) due to the dependence on initial and lateral BCs from an ultimately global model, modify these results. The resulting idealized ensembles show high variability of time and position of convective triggering, as well as multiple convective modes. The

orographic forcing remains visible in the ensembles, but is smaller and hardly quantifiable. In those situations, not the position of convection is predicted, but the large-scale conditions and their influence on the emergence and development of convective cells.

The OSSE in this study exhibits no model errors or systematic biases. Therefore, the predictability of the model state and the model predictability of the atmospheric state are identical. Hence, we analyzed short-term forecasts over Germany with advanced radar DA to overcome this shortcoming and add to our understanding. The 6 h forecasts in a domain over Germany during a High Impact Weather (HIW) period that displays a meteorological situation comparable to the idealized experiments is a next step to generalize our findings so far. The HIW period is especially interesting as it features several extreme convective precipitation events, which are intrinsically less predictable (Sterk et al., 2016). As those experiments are evaluated with radar-derived observations, model errors and representativity issues are introduced.

In our study, the orographically more influenced South of Germany shows significantly higher skill throughout the forecasts than the comparatively plain North measured by the decorrelation scale, and the difference is increasing with lead time, which corroborates increased predictability of the model state of deep convection in Southern Germany due to orography. In other words, a state-of-the-art model produces less variability in the location of deep convection in the ensemble forecast. In a probabilistic sense, convective precipitation over orographic regions is more constrained to those regions favored by the interaction of orography and, for example, synoptic-scale flow or radiation, increasing the predictability. Higher predictability of the model state does not necessarily result in an improved weather forecast, the reason being that the model might miss or misrepresent relevant processes in the atmosphere that are unresolved and parameterized.

The application of the believable scale assesses the model predictability of the atmospheric state or the skill of the forecasts, and the comparison to the decorrelation scale represents an estimate of the model error. We identify higher model predictability of the atmospheric state in the more orographic South. This finding is especially striking, as previous studies found that interactions between the synoptic scales and orography (Nuss and Miller, 2001; Picard and Mass, 2017), or unresolved processes related to orography, like convergence-lines or radiation on slopes (Barthlott et al., 2011), pose additional challenges to NWP systems. However, the findings of the current study do not support the previous research. Instead, our results suggest that the COSMO-KENDA model can cope with those additional challenges to a degree that allows us to transfer the predictability provided by nature to the forecast. The difference between the two predictability metrics hints towards a miss- or underrepresentation of convection triggers in Northern Germany in the model. The fact that the pre-operational model version with direct assimilation of radar observations is able to project predictability of the model state to the atmospheric state is very encouraging, as it highlights the potential of current model versions.

Our attempt to verify those findings in COSMO-DE-EPS, the NWP system operational between 2014 and 2016 at DWD, delivers more ambiguous results. We confirm the

findings regarding increased predictability of the model state in Southern Germany, but could not detect increased model predictability of the atmospheric state. The overall skill in this model configuration is, in contrast to the pre-operational one, not sufficient to see the beneficial impact of orography. The performance of the forecast initialized at 1200 UTC corroborates the fact that COSMO-DE-EPS is unable to predict the position of convection for lead times above 3 h. For the forecasts initialized at 0000 UTC, we additionally have to acknowledge the lead time of at least 12 h at the beginning of the convectively most active period, which is beyond the predictability estimates for convective precipitation (e.g. Schwartz et al., 2015, or this study).

Based on the convective adjustment timescale, we divided the three years of data into two categories, which allow differentiating local and synoptic forcing situations (Done et al., 2006; Keil and Craig, 2011; Done et al., 2012). We found that the scale on which COSMO-DE-EPS forecasts are reasonably skillful grows within just a few hours to a constant value that is defined by the synoptic situation. Local forcing situations depict a believable scale of $\mathcal{O}(300 \text{ km})$, independent of the region of Germany. Interestingly, Surcel et al. (2015) also report 300 km in precipitation forecasts for the United States without radar DA. In this situation, the influence of the weather regime prevails over the possible impact of orography. Situations with synoptic forcing show a believable scale of $\mathcal{O}(100 \text{ km})$, which is in the same order as our idealized experiments and comparable to previous studies (Schwartz et al., 2009; Clark et al., 2010; Mittermaier et al., 2013), based on FSS or other neighborhood methods. Done et al. (2006, 2012) found that those situations exhibit only predictable area-averaged precipitation, but low predictability in terms of location.

In conclusion, our study concurs with previous literature that synoptically forced situations exhibit increased predictability of convection compared to locally forced ones (Done et al., 2006; Keil and Craig, 2011; Done et al., 2012), which is another potential source of predictability mentioned by Anthes (1986). Interestingly, we found the most considerable beneficial impact of orography in the late afternoon and local forcing regime, which represents a convectively active situation that is primarily dependent on a surface trigger like orography, highlighting the interconnection of the weather regime and orography. Recent research at the ECMWF already tries to exploit the higher predictability of favorable conditions for extreme convection related to atmospheric rivers compared to the actual predictability of these events. Lavers et al. (2014, 2016) show that medium-range forecasts can be skillful in this regard for almost two weeks. However, this extended forecast horizon results in a higher false alarm rate and is different in the sense that it is not concerned with the forecast of a single convective cell, but overall higher predictability of extreme precipitation.

4.2. Radar DA and Orography

The impact of radar DA is an essential extension of the results presented so far. Several operational centers, like DWD, plan operational 3D radar DA within 2019 and we

believe that the assimilation of radar observations can provide the NWP system with similar information as orography—the initial position of the convective cell. Additionally, realistic ICs play an essential role in studies of practical predictability, and therefore studies based solely on idealized, homogeneous profiles might have limited explanatory power concerning forecast capabilities (Weyn and Durran, 2017).

We find that radar DA has a profoundly beneficial impact on the practical predictability in idealized experiments that disregard observation uncertainties such as representativeness issues or systematic and correlated errors by using a perfect model approach creating spatio-temporally highly resolved *synthetic* observations from a Nature Run, and thereby neglecting approximations and inconsistencies in the description of atmospheric processes in the NWP model or the forward operator. The convection is correctly positioned after the DA cycling, and the forecast ensembles generally develop in good agreement with the Nature Run. The forecast horizon is extended by 6 h in experiments that account only for IC uncertainty, but not for synoptic-scale uncertainty. According to our expectations, the assimilation of radar observation provides the forecasts with information on the initial position of convective cells, which can be propagated correctly in a perfect model environment. Radar DA compensates the beneficial impact of orography for around 4 h in this configuration.

The introduction of additional realistic, synoptic-scale IC uncertainty decreases the overall skill of the forecast, but the corrections by radar DA initially compensate for some of the synoptic-scale uncertainties. However, the analysis represents the higher uncertainty of the first guess ensemble with a higher initial spread on all scales. Due to that, the error growth and consequently the loss of predictability on ever-growing scales accelerates in those experiments. The forecast horizon, however, is again extended by 6 h, while the improvement in predictable horizontal scales is even larger. We find that in the presence of only IC uncertainty, scales of around 40 km remain predictable after 6 h lead time. To put this into perspective, in the presence of realistic IC and LBC uncertainty scales below 100 km become unpredictable. Radar DA proves its ability to correct flawed, synoptic-scale conditions in this setup, further emphasizing the potential of directly assimilating radar observations and demonstrating that an LETKF system is capable of effectively using reflectivity observations.

The idealized and the pre-operational, short-range COSMO forecasts in the HIW period agree remarkably well on the scales that remain predictable at certain lead times ($\mathcal{O}(100\text{ km})$ after 6 h). The forecast horizon after which predictability on convective scales due to IC is lost amounts to about 6 h in the idealized and 3 h in the more realistic COSMO forecasts. It thus highlights both a further potential for improvement in a full NWP model including model errors, as well as realistic initial and boundary conditions, but also the already high performance of radar DA at this point. Zhang et al. (2016), for example, found an intrinsic predictability limit of 3–6 h for a tornadic thunderstorm event. It is interesting that the LHN that was at this point operational, and the direct 3D assimilation of radar DA show comparable predictive skill. It is above all important to introduce the information about convection in the system regularly and in real-time, while the method to do so is of minor importance.

The major differences of this idealized setup to a real NWP system are the absence of systematic deficiencies in the model, approximations in the observation operator, and

uncertainty in the observations. These impacts can be estimated by comparison of the decorrelation and the believable scale. While the predictability of the model state lies around 50 km after 6 h, the model predictability of the atmospheric state dropped to 150 km for Germany. Therefore, we postulate that mitigating such deficiencies is of utmost importance for the assimilation of radar observations in real systems. We also see sensitivities that arise from the interplay of model parameters, like the mixing length, and the choice of the DA algorithm, underlining the importance of joint development of both parts of a NWP model. The comparison of the forecast skills of COSMO-DE-EPS and COSMO-KENDA implicate that this approach has already been successful in the past.

Finally, our approach allows for the identification of further important sensitivities of the ensemble DA system KENDA in the idealized OSSE environment. The magnitude of the assigned observation error has a major impact on the predictive skill. Increasing the observation error of radar reflectivity and velocity by a factor of three (amounting to 15 dBZ and 3 m s^{-1}) increases, i.e. worsens, the skillful scales by 20 km to 40 km, depending on the lead time. Nevertheless, the radar DA, using fairly large assigned observation errors, still improves the predictive scales compared to experiments without DA. In contrast, smaller observation errors can further extend the forecast horizon, but the gain becomes smaller and, in the end, infinitesimal. Our experiments reveal that we are not yet approaching this intrinsic limit of predictability on the convective scale.

Interestingly, the assimilation of only velocity or reflectivity information leads to skill close to the benchmark experiment, where both velocity and reflectivity observations are assimilated. This encourages efforts to enlarge the amount of available radar velocity information (e.g., from clear-air returns) as in operational systems, velocity observations are usually easier to assimilate than reflectivity. Radar forward operators that include reflectivity are complicated due to non-linearity of the measurements and the fact that neither reflectivity nor precipitation are prognostic variables of the model. All sensitivity experiments consistently outperform the Background ensemble without DA for the lead time of 6 h.

In the sensitivity experiments, the impact of orography is more significant and earlier recognizable, indicating that orography as a lower boundary forcing can potentially correct flawed ICs (as suggested by Laprise et al., 2000) or observations, and provides a source of predictability for convection. In reality, the increased predictability due to dynamical interactions of the flow with the orography could be partly compensated by incomplete parametrizations of microphysics or turbulence. Those have, for instance, a considerable impact on the latent heat release in the thunderstorm and are therefore crucial for its development. However, the comparison to the pre-operational forecasts, which suffer from those limitations, highlights that some of the beneficial impact of orography is replicated in the presence of complex terrain and model errors. This further emphasizes the potential of convection-allowing EPS building on initial states derived from sophisticated DA systems.

Furthermore, we looked into the predictability of convection based on their level of organization and the convective mode. The choice of the DA time window with respect

to the stage of the convective life cycle is essential and our results indicate a reduction of the forecast horizon by up to 2 h when choosing the DA window at a later stage of the convective life cycle, simply because the assimilated convective cells cease to exist sooner.

We explored the predictability and model performance in two different convective situations or modes. One depicts almost evenly distributed small convective cells, while a large, well-developed MCS dominates the other one. We find that the predictability of the model state is comparable in both situations, but the ensemble of the unorganized case diverges much faster from the respective Nature Run. We call this a reduced ‘model predictability of the analysis state’. By this, we mean that although the analyses are comparably skillful in the beginning, the forecast model can propagate the organized one better, which we attribute to a dynamically more balanced analysis (Lange, 2016). In conclusion, practical predictability is a function of the whole NWP system, consisting of the combination of model and data assimilation, and increases for well-organized convection at this point.

The results in both model configurations highlight that the assimilation of radar data has the potential for drastic improvements in the predictive skill of convective precipitation. Under realistic conditions, the assimilation of radar data potentially has a long-lasting impact and increases the forecast horizon by up to 6 h. Pre-operational forecasts already improve for up to 3 h. We are aware that, e.g. a different convective environment, the degree of convective organization, a different shape or height of the orography, the magnitude of synoptic-scale errors, surface and atmospheric heterogeneities, as well as model error, will affect the exact numbers presented in this study. However, we believe that the relative magnitude of the various ensemble forecasting experiments is representative of state-of-the-art NWP.

4.3. Outlook on Future Applications

The combination of decorrelation and believable scale assesses different facets of practical predictability conveniently, as both scores provide a predictable scale, an accessible measure with direct physical meaning. We believe the application and combination of both metrics is a novel way to shed further light not only in predictability research, but also in the assessment of model performance. As the decorrelation scale is calculated exclusively from ensemble information, it can provide forecasters with live guidance on forecast uncertainty on the convective scale. It can also be used to remove details on unpredictable scales, which simplifies the visual interpretation of forecasts for professionals and end-users. DWD plans to implement such a feature in WarnWetter, its official weather warning application. Both metrics are also included in the python package ‘enstools’ developed in Waves to Weather (W2W), which will allow the research community easy access to those metrics soon.

The quantification of forecast uncertainty is an essential part of every forecast. This study shows the spatial and temporal characteristics of practical predictability limits. The existence of improved predictability in the vicinity of orography and in certain weather regimes, and their interconnection is valuable information and can potentially enhance the prediction of forecast uncertainty. A forecast improved in this regard will promote decision making for the general public and build trust. But especially emergency services demand probabilistic information for mission planning. Hence this work not only enables advancements in weather forecasting, but also creates a toolbox for research applications.

The OSSE, developed by my colleague Heiner Lange (Lange, 2016) and me, proved to be a useful tool for a wide range of research and triggered the development of a similar system at the Deutscher Wetterdienst (DWD) and will also be used for further studies on satellite data assimilation at the Meteorologisches Institut München (MIM).

Moreover, the idealized OSSE framework provides an attractive experimental design to look into the predictability of different convective modes or convection with a higher or lower level of organization. The advantage is that those convective states develop from similar initial conditions rendering the surrounding atmospheric conditions comparable, thereby isolating the effect of the convective organization in a novel way. These background conditions can be modified to develop more extreme events with inherently lower intrinsic predictability, such as severe thunderstorms on the verge of producing tornadoes. The specific predictability limits of those storms and their sensitivities to the environment are of immense practical importance. Furthermore, previous studies tie the occurrence of convective modes with tornadoes to orographic features, linking this line of research with another aspect of this dissertation.

A List of Abbreviations

NWP	Numerical Weather Prediction
ENIAC	Electronic Numerical Integrator and Computer
DWD	Deutscher Wetterdienst
SINFONY	Seamless INtegrated FOrcastiNg sYstem
ECMWF	European Centre for Medium-Range Weather Forecasts
NCEP	National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
MIM	Meteorologisches Institut München
W2W	Waves to Weather
EPS	Ensemble Prediction System
LAM	Limited-Area Model
OSSE	Observing System Simulation Experiment
HWT	Hazardous Weather Testbed
COSMO	Consortium for small-scale modeling model
ICON-EU	Icosahedral Nonhydrostatic Model over European Union
GME	Global Model Erde
IFS	Integrated Forecasting System
GFS	Global Forecasting System
DA	Data Assimilation
KENDA	Kilometre-Scale Ensemble Data Assimilation
EnKF	Ensemble Kalman Filter

LETKF	local ensemble transform Kalman filter
LHN	latent heat nudging
3DVar	Three Dimensional Variational Data Assimilation
4DVar	Four Dimensional Variational Data Assimilation
EMVORADO	efficient modular volume-scanning radar forward operator
PAR	phased-array radar
TERRA-ML	multi-layer soil model
IC	initial condition
BC	boundary condition
LBC	lateral boundary condition
HIW	High Impact Weather
MAP	Mesoscale Alpine Programme
COPS	Convective and Orographically-induced Precipitation Study
CAPE	Convective Available Potential Energy
CIN	Convective Inhibition
MCS	Mesoscale Convective System
MSE	Mean Square Error
RMSE	Root Mean Squared Error
NRMSE	Normalized Root Mean Square Error
FSS	Fractions Skill Score
SAL	Structure, Amplitude and Location Score
DAS	Displacement and Amplitude Score
FFT	Fast Fourier Transformation
DCT	Discrete Cosine Transformation

B List of Experiments

Table 1: Summary of abbreviations used to describe the experiments.

Abbreviation	Description
<i>flat</i>	Flat experiment without orography
<i>oro</i>	Orography experiment with Gaussian-shaped mountain of 1000 m height
<i>well-organized</i>	experiment based on ensemble member with organized convection
<i>popcorn-like</i>	experiment based on ensemble member with popcorn-like convection
IC	only initial condition errors
IBC	initial and additional boundary condition errors
IDEAL-Nat	deterministic Nature Run representing the truth, i.e. providing observations
IDEAL	idealized COSMO-KENDA ensemble forecast without any data assimilation
IDEAL-DA	idealized COSMO-KENDA ensemble forecast started from analysis using velocity and reflectivity observations in DA
IDEAL-DAtime	Ensemble forecast started from assimilation window, time-shifted by 4 h to 1100 UTC – 1200 UTC
IDEAL-DAobserr05	Ensemble with halved assigned observation errors for reflectivity and velocity (to 2.5 dBZ and 0.5 m s^{-1})
IDEAL-DAobserr3	Ensemble with tripled assigned observation errors for reflectivity and velocity (to 15 dBZ and 3 m s^{-1})
IDEAL-DAwind	Ensemble forecast from analysis using only velocity observations in DA
IDEAL-DArelectivity	Ensemble forecast from analysis using only reflectivity observations in DA
DE-DA-LHN	COSMO-KENDA ensemble with latent heat nudging valid for a high impact weather period of 14 days
DE-DA-3DRad	COSMO-KENDA ensemble with direct assimilation of 3D reflectivities and radial velocity and warm bubbles triggered by analysis of missing reflectivities in the model valid for a high impact weather period of 14 days
DE-EPS	COSMO-DE-EPS ensemble over 3 summer periods between 2014-2016

List of Figures

1.1	Schematic diagram visualizing the ensemble approach to NWP	3
1.2	Schematic of practical versus intrinsic predictability	4
1.3	Visualization of the Lorenz attractor	6
1.4	Scale-dependent predictability limits	7
2.1	Conceptual visualization of a Local Ensemble Transform Kalman Filter	21
2.2	Sounding of the idealized experiments	24
2.3	Flow diagram of experimental setup in idealized experiments	26
2.4	Flow diagram of the cycling in the HIW period:	29
2.5	Map of investigated Model Domain over Germany	30
2.6	Plot of the Ratio used to find the decorrelation scale	34
2.7	Illustration of the verification with a neighborhood approach	35
2.8	Illustration of the verification testbed	37
2.9	Verification of forecasts with location errors with different metrics . . .	38
2.10	Verification of forecasts with amplitude and structure errors with different metrics	39
3.1	Development of IDEAL-Nat	43
3.2	Time series of the precipitation for IDEAL	44
3.3	Spaghetti plots of IDEAL at 0800 UTC	45
3.4	Spaghetti plots of IDEAL at 1000 UTC	46
3.5	Spaghetti plots of IDEAL at 1200 UTC	47

3.6	Snapshot of 5 representative ensemble members for IDEAL IC <i>flat</i> and IDEAL IBC <i>flat</i>	48
3.7	Snapshot of 5 representative ensemble members for IDEAL IC <i>oro</i> and IDEAL IBC <i>oro</i>	49
3.8	Time series of NRMSE for IDEAL IC	50
3.9	Time series of the NRMSE for IDEAL IBC and IDEAL-DA IBC	51
3.10	FSS tables for the IDEAL experiments	52
3.11	Time series of the decorrelation and believable scale of IDEAL	53
3.12	Spaghetti plots of IDEAL-DA at 0800 UTC	57
3.13	Spaghetti plots of IDEAL-DA at 1000 UTC	58
3.14	Spaghetti plots of IDEAL-DA at 1200 UTC	59
3.15	FSS tables for the IDEAL-DA experiments	60
3.16	Time series of the decorrelation and believable scale for IDEAL-DA	61
3.17	Time series of the decorrelation and believable scale for IDEAL-DA observation error sensitivity experiments	63
3.18	Time series of the decorrelation and believable scale for IDEAL-DA observation type sensitivity experiments	64
3.19	Time series of the decorrelation and believable scale for IDEAL-DA time sensitivity experiments	65
3.20	Stamp plot of IDEAL IBC <i>oro</i> at 0830 UTC	67
3.21	Development of IDEAL-Nat <i>convective mode</i>	68
3.22	Spaghetti plots of IDEAL-DA <i>convective mode</i> at 0800 and 1000 UTC	69
3.23	Time series of the number of convective cells for IDEAL-DA <i>convective mode</i>	70
3.24	Spectra for IDEAL-DA <i>convective mode</i>	70
3.25	Time series of the decorrelation and believable scale for IDEAL-DA <i>convective mode</i>	71
3.26	Time series of the decorrelation and believable scale for COSMO-KENDA in the HIW period	74
3.27	Time series of NRMSE for COSMO-DE-EPS	77
3.28	Time series of the decorrelation and believable scale for COSMO-DE-EPS	78

Bibliography

- Abbe, C., 1901: The Physical Basis of Long-Range Weather Forecasts. *Monthly Weather Review*, **29** (29), 551–561,
- Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert, 2009: Application of Spatial Verification Methods to Idealized and NWP-Gridded Precipitation Forecasts. *Weather and Forecasting*, **24** (6), 1485–1497,
- Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses. *Monthly Weather Review*, **137** (6), 1805–1824,
- Aksoy, A., D. C. Dowell, and C. Snyder, 2010: A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part II: Short-Range Ensemble Forecasts. *Monthly Weather Review*, **138** (4), 1273–1292,
- Anthes, R. A., 1986: The General Question of Predictability. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., Vol. II, American Meteorological Society, Boston, MA, 636–656,
- Ashley, S. T., and W. S. Ashley, 2008: Flood Fatalities in the United States. *Journal of Applied Meteorology and Climatology*, **47** (3), 805–818,
- Bachmann, K., C. Keil, G. C. Craig, M. Weissmann, and C. Welzbacher, 2019a: Predictability of Deep Convection in Idealized and Operational Forecasts : Effects of Radar Data Assimilation , Orography and Synoptic Weather Regime.
- Bachmann, K., C. Keil, and M. Weissmann, 2019b: Impact of radar data assimilation and orography on predictability of deep convection. *Quarterly Journal of the Royal Meteorological Society*, **145** (718), qj.3412,
- Baldauf, M., G. Doms, U. Schättler, C. Schraff, and U. Blahak, 2018:
- Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Monthly Weather Review*, **139** (12), 3887–3905,
- Barrett, A. I., S. L. Gray, D. J. Kirshbaum, N. M. Roberts, D. M. Schultz, and J. G. Fairman, 2015: Synoptic versus orographic control on stationary convective banding. *Quarterly Journal of the Royal Meteorological Society*, **141** (689), 1101–1113,

- Barthlott, C., and Coauthors, 2011: Initiation of deep convection at marginal instability in an ensemble of mesoscale models: A case-study from COPS. *Quarterly Journal of the Royal Meteorological Society*, **137** (SUPPL. 1), 118–136,
- Bauer, P., A. J. Geer, P. Lopez, and D. Salmond, 2010: Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation. *Quarterly Journal of the Royal Meteorological Society*, **136** (652), 1868–1885,
- Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525** (7567), 47–55,
- Baur, F., C. Keil, and G. C. Craig, 2018: Soil moistureprecipitation coupling over Central Europe: Interactions between surface anomalies at different scales and the dynamical implication. *Quarterly Journal of the Royal Meteorological Society*, **144** (717), 2863–2875,
- Ben Bouallègue, Z., and S. E. Theis, 2014: Spatial techniques applied to precipitation ensemble forecasts: from verification results to probabilistic products. *Meteorological Applications*, **21** (4), 922–929,
- Berri, G. J., and J. Paegle, 1990: Sensitivity of Local Predictions to Initial Conditions. *Journal of Applied Meteorology*, **29** (3), 256–267,
- Bick, T., and Coauthors, 2016: Assimilation of 3D Radar Reflectivities with an Ensemble Kalman Filter on the Convective Scale. *submitted to Quarterly Journal of the Royal Meteorological Society*, **142** (1990), 1–23,
- Bierdel, L., C. Snyder, S.-H. Park, and W. C. Skamarock, 2016: Accuracy of Rotational and Divergent Kinetic Energy Spectra Diagnosed from Flight-Track Winds. *Journal of the Atmospheric Sciences*, **73** (8), 3273–3286,
- Bischof, M., 2011: Ensemble Simulations of Convective Storms. Ph.D. thesis,
- Bjerknes, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteorologische Zeitschrift*,
- Bormann, N., M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally, 2016: Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, **142** (697), 1767–1780,
- Bormann, N., S. Saarinen, G. Kelly, and J.-N. Thépaut, 2003: The Spatial Structure of Observation Errors in Atmospheric Motion Vectors from Geostationary Satellite Data. *Monthly Weather Review*, **131** (4), 706–718,
- Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution Requirements for the Simulation of Deep Moist Convection. *Monthly Weather Review*, **131** (10), 2394–2416,
- Campbell, W. F., E. A. Satterfield, B. Ruston, and N. L. Baker, 2017: Accounting for Correlated Observation Error in a Dual-Formulation 4D Variational Data Assimilation System. *Monthly Weather Review*, **145** (3), 1019–1032,
- Carbone, R. E., and J. D. Tuttle, 2008: Rainfall occurrence in the U.S. warm season: The diurnal cycle. *Journal of Climate*, **21** (16), 4132–4146,

- Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, **11** (2), 141–154,
- Charney, J. G., R. Fjörtoft, and J. V. Neumann, 1950: Numerical Integration of the Barotropic Vorticity Equation. *Tellus*, **2** (4), 237–254,
- Cintineo, R. M., J. A. Otkin, T. A. Jones, S. E. Koch, and D. J. Stensrud, 2016: Assimilation of Synthetic GOES-R ABI Infrared Brightness Temperatures and WSR-88D Radar Observations in a High-Resolution OSSE. *Monthly Weather Review*, **144** (9), 3159–3180,
- Cintineo, R. M., and D. J. Stensrud, 2013: On the Predictability of Supercell Thunderstorm Evolution. *Journal of the Atmospheric Sciences*, **70** (7), 1993–2011,
- Clark, A. J., W. A. Gallus, and M. L. Weisman, 2010: Neighborhood-Based Verification of Precipitation Forecasts from Convection-Allowing NCAR WRF Model Simulations and the Operational NAM. *Weather and Forecasting*, **25** (5), 1495–1509,
- Clark, P. A., N. M. Roberts, H. W. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, **23** (2), 165–181,
- Corsmeier, U., and Coauthors, 2011: Processes driving deep convection over complex terrain: a multi-scale analysis of observations from COPS IOP 9c. *Quarterly Journal of the Royal Meteorological Society*, **137** (S1), 137–155,
- Craig, G. C., C. Keil, and D. Leuenberger, 2012: Constraints on the impact of radar rainfall data assimilation on forecasts of cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, **138** (663), 340–352,
- Cronin, T. W., K. A. Emanuel, and P. Molnar, 2015: Island precipitation enhancement and the diurnal cycle in radiative-convective equilibrium. *Quarterly Journal of the Royal Meteorological Society*, **141** (689), 1017–1034,
- Demko, J. C., and B. Geerts, 2010: A Numerical Study of the Evolving Convective Boundary Layer and Orographic Circulation around the Santa Catalina Mountains in Arizona. Part II: Interaction with Deep Convection. *Monthly Weather Review*, **138** (5), 3603–3622,
- Denis, B., J. Côté, and R. Laprise, 2002: Spectral Decomposition of Two-Dimensional Atmospheric Fields on Limited-Area Domains Using the Discrete Cosine Transform (DCT). *Monthly Weather Review*, **130** (7), 1812–1829,
- Dey, S. R. A., G. Leoncini, N. M. Roberts, R. S. Plant, and S. Migliorini, 2014: A Spatial View of Ensemble Spread in Convection Permitting Ensembles. *Monthly Weather Review*, **142** (11), 4091–4107,
- Dey, S. R. A., R. S. Plant, N. M. Roberts, and S. Migliorini, 2016: Assessing spatial precipitation uncertainties in a convective-scale ensemble. *Quarterly Journal of the Royal Meteorological Society*, **142** (701), 2935–2948, qj.2893.
- Dixon, M., Z. Li, H. W. Lean, N. M. Roberts, and S. Ballard, 2009: Impact of Data Assimilation on Forecasting Convection over the United Kingdom Using a High-Resolution Version of the Met Office Unified Model. *Monthly Weather Review*, **137** (5), 1562–1584,

- Done, J., G. C. Craig, S. L. Gray, and P. A. Clark, 2012: Case-to-case variability of predictability of deep convection in a mesoscale model. *Quarterly Journal of the Royal Meteorological Society*, **138** (664), 638–648,
- Done, J., G. C. Craig, S. L. Gray, P. A. Clark, and M. E. B. Gray, 2006: Mesoscale simulations of organized convection: Importance of convective equilibrium. *Quarterly Journal of the Royal Meteorological Society*, **132** (616), 737–756,
- Done, J., C. A. Davis, and M. L. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmospheric Science Letters*, **5** (6), 110–117,
- Duda, J. D., and W. A. Gallus, 2013: The Impact of Large-Scale Forcing on Skill of Simulated Convective Initiation and Upscale Evolution with Convection-Allowing Grid Spacings in the WRF*. *Weather and Forecasting*, **28** (4), 994–1018,
- Durrán, D. R., and M. Gingrich, 2014: Atmospheric Predictability: Why Butterflies Are Not of Practical Importance. *Journal of the Atmospheric Sciences*, **71** (7), 2476–2488,
- Durrán, D. R., and J. A. Weyn, 2016: Thunderstorms Do Not Get Butterflies. *Bulletin of the American Meteorological Society*, **97** (2), 237–243,
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99** (C5), 10 143,
- Evensen, G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53** (4), 343–367,
- Faggian, N., B. Roux, P. Steinle, and E. E. Ebert, 2014: Fast calculation of the Fraction Skill Score. *Mausam*,
- Flack, D. L. A., S. L. Gray, R. S. Plant, H. W. Lean, and G. C. Craig, 2018: Convective-Scale Perturbation Growth across the Spectrum of Convective Regimes. *Monthly Weather Review*, **146** (1), 387–405,
- Flack, D. L. A., R. S. Plant, S. L. Gray, H. W. Lean, C. Keil, and G. C. Craig, 2016: Characterisation of convective regimes over the British Isles. *Quarterly Journal of the Royal Meteorological Society*, **142** (696), 1541–1553,
- Foresti, L., I. V. Sideris, L. Panziera, D. Nerini, and U. Germann, 2018: A 10-year radar-based analysis of orographic precipitation growth and decay patterns over the Swiss Alpine region. *Quarterly Journal of the Royal Meteorological Society*, **144** (716), 2277–2301,
- Gandin, L. S., 1966: Objective analysis of meteorological fields. By L. S. Gandin. Translated from the Russian. Jerusalem (Israel Program for Scientific Translations), 1965. Pp. vi, 242: 53 Figures; 28 Tables. \$4 1s. 0d. *Quarterly Journal of the Royal Meteorological Society*, **92** (393), 447–447,
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125** (554), 723–757,
- Gastaldo, T., and Coauthors, 2018: Data assimilation of radar reflectivity volumes in a LETKF scheme. *Nonlinear Processes in Geophysics*, **25** (4), 747–764,

- Gebhardt, C., S. Theis, P. Krahe, and V. Renner, 2008: Experimental ensemble forecasts of precipitation based on a convection-resolving model. *Atmospheric Science Letters*, **9** (2), 67–72,
- Gebhardt, C., S. E. Theis, M. Paulat, and Z. Ben Bouallègue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmospheric Research*, **100** (2-3), 168–177,
- Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, **24** (5), 1416–1430,
- Gneiting, T., and M. Katzfuss, 2014: Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, **1** (1), 125–151,
- Gneiting, T., and A. E. Raftery, 2007: Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102** (477), 359–378,
- Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2018: Validation of Mountain Precipitation Forecasts from the Convection-Permitting NCAR Ensemble and Operational Forecast Systems over the Western United States. *Weather and Forecasting*, (July),
- Gustafsson, N., and Coauthors, 2017: Survey of data assimilation methods for convective scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, **24** (0000), 1–24,
- Hamill, T. M., and J. S. Whitaker, 2011: What Constrains Spread Growth in Forecasts Initialized from Ensemble Kalman Filters? *Monthly Weather Review*, **139** (1), 117–131,
- Hanley, K. E., R. S. Plant, T. H. Stein, R. J. Hogan, J. C. Nicol, H. W. Lean, C. Halliwell, and P. A. Clark, 2015: Mixing-length controls on high-resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society*, **141** (686), 272–284,
- Harnisch, F., and C. Keil, 2015: Initial Conditions for Convective-Scale Ensemble Forecasting Provided by Ensemble Data Assimilation. *Monthly Weather Review*, **143** (5), 1583–1600,
- Hassanzadeh, H., J. Schmidli, W. Langhans, L. Schlemmer, and C. Schär, 2014: Impact of topography on the diurnal cycle of summertime moist convection in idealized simulations. *Meteorologische Zeitschrift*, **25** (Dc), 181–194,
- Hirt, M., S. Rasp, U. Blahak, and G. C. Craig, 2019: Process based stochastic perturbations for convection initiation in km-scale models.
- Hohenegger, C., D. Lüthi, and C. Schär, 2006: Predictability Mysteries in Cloud-Resolving Models. *Monthly Weather Review*, **134** (8), 2095–2107,
- Hohenegger, C., and C. Schär, 2007a: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bulletin of the American Meteorological Society*, **88** (11), 1783–1793,
- Hohenegger, C., and C. Schär, 2007b: Predictability and Error Growth Dynamics in Cloud-Resolving Models. *Journal of the Atmospheric Sciences*, **64** (12), 4467–4478,
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Monthly Weather Review*, **133** (3), 604–620,

- Houtekamer, P. L., and F. Zhang, 2016: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **144** (12), MWR-D-15-0440.1,
- Houze, R. A., 2012: Orographic effects on precipitating clouds. *Reviews of Geophysics*, **50** (1), RG1001,
- Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, **230** (1-2), 112–126, 0511236.
- Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. *Tellus A: Dynamic Meteorology and Oceanography*, **56** (4), 273–277,
- Isotta, F. A., and Coauthors, 2014: The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, **34** (5), 1657–1675,
- Johnson, A., X. Wang, J. R. Carley, L. J. Wicker, and C. D. Karstens, 2015: A Comparison of Multiscale GSI-Based EnKF and 3DVar Data Assimilation Using Radar and Conventional Observations for Midlatitude Convective-Scale Precipitation Forecasts. *Monthly Weather Review*, **143** (8), 3087–3108,
- Johnson, A., and Coauthors, 2014:
- Jones, T. A., K. Knopfmeier, D. M. Wheatley, G. J. Creager, P. Minnis, and R. Palikonda, 2016: Storm-Scale Data Assimilation and Ensemble Forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined Radar and Satellite Data Experiments. *Weather and Forecasting*, **31** (1), 297–327,
- Judt, F., 2018: Insights into Atmospheric Predictability Through Global Convection-Permitting Model Simulations. *Journal of the Atmospheric Sciences*, **75** (5), 1477–1497,
- Kain, J. S., and Coauthors, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bulletin of the American Meteorological Society*, **94** (8), 1213–1225,
- Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, **82** (1), 35,
- Katzfuss, M., J. R. Stroud, and C. K. Wikle, 2016: Understanding the Ensemble Kalman Filter. *The American Statistician*, **70** (4), 350–357,
- Keil, C., and G. C. Craig, 2009: A Displacement and Amplitude Score Employing an Optical Flow Technique. *Weather and Forecasting*, **24** (2008), 1297–1308,
- Keil, C., and G. C. Craig, 2011: Regime-dependent forecast uncertainty of convective precipitation. *Meteorologische Zeitschrift*, **20** (2), 145–151,
- Keil, C., F. Heinlein, and G. C. Craig, 2014: The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quarterly Journal of the Royal Meteorological Society*, **140** (679), 480–490,
- Kirshbaum, D. J., 2011: Cloud-Resolving Simulations of Deep Convection over a Heated Mountain. *Journal of the Atmospheric Sciences*, **68** (2), 361–378,

- Kovacs, M., and D. J. Kirshbaum, 2016: Topographic Impacts on the Spatial Distribution of Deep Convection over Southern Quebec. *Journal of Applied Meteorology and Climatology*, **55** (3), 743–762,
- Kühnlein, C., C. Keil, G. C. Craig, and C. Gebhardt, 2014: The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Quarterly Journal of the Royal Meteorological Society*, **140** (July), 1552–1562,
- Kuo, J.-T., and H. D. Orville, 1973: A Radar Climatology of Summertime Convective Clouds in the Black Hills. *Journal of Applied Meteorology*, **12** (2), 359–368,
- Lange, H., 2016: On the Use of Radar and Aircraft Data in Ensemble Data Assimilation of Convection for Non-hydrostatic Numerical Weather Prediction. Ph.D. thesis,
- Lange, H., and G. C. Craig, 2014: The Impact of Data Assimilation Length Scales on Analysis and Prediction of Convective Storms. *Monthly Weather Review*, **142** (10), 140707140511 002,
- Lange, H., G. C. Craig, and T. Janjić, 2017: Characterizing noise and spurious convection in convective data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143** (709), 3060–3069,
- Lange, H., and T. Janjić, 2016: Assimilation of Mode-S EHS Aircraft Observations in COSMO-KENDA. *Monthly Weather Review*, **144** (5), 1697–1711,
- Laplace, P.-S., 2012: *Pierre-Simon Laplace Philosophical Essay on Probabilities: Translated from the fifth French edition of 1825 With Notes by the Translator*.
- Laprise, R., M. R. M. Varma, B. Denis, D. Caya, I. Zawadzki, . C. i. B. D. Sciences, and the Earth, 2000: Predictability of a Nested Limited-Area Model. *Monthly Weather Review*, **128** (12), 4149–4154,
- Lavers, D. A., F. Pappenberger, D. S. Richardson, and E. Zsoter, 2016: ECMWF Extreme Forecast Index for water vapor transport: A forecast tool for atmospheric rivers and extreme precipitation. *Geophysical Research Letters*, **43** (22), 11,852–11,858, 2014WR016527.
- Lavers, D. A., F. Pappenberger, and E. Zsoter, 2014: Extending medium-range predictability of extreme hydrological events in Europe. *Nature Communications*, **5** (1), 5382,
- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *Journal of Computational Physics*, **227** (7), 3515–3539,
- Levizzani, V., F. Pinelli, M. Pasqui, S. Melani, A. G. Laing, and R. E. Carbone, 2010: A 10-year climatology of warm-season cloud patterns over Europe and the Mediterranean from Meteosat IR observations. *Atmospheric Research*, **97** (4), 555–576,
- Lilly, D. K., 1990: Numerical prediction of thunderstormshas its time come? *Quarterly Journal of the Royal Meteorological Society*, **116** (494), 779–798,
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk Parameterization of the Snow Field in a Cloud Model. *Journal of Climate and Applied Meteorology*, **22** (6), 1065–1092,
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **112** (474), 1177–1194,

- Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, **20** (2), 130–141,
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21** (3), 289–307,
- Lorenz, E. N., 1996: Predictability a problem partly solved. *Predictability of Weather and Climate*, T. Palmer, and R. Hagedorn, Eds., Cambridge University Press, Cambridge, 40–58,
- Lynch, P., 2008: The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, **227** (7), 3431–3444,
- Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. *Bulletin of the American Meteorological Society*,
- Melhauser, C., and F. Zhang, 2012: Practical and Intrinsic Predictability of Severe and Convective Weather at the Mesoscales. *Journal of the Atmospheric Sciences*, **69** (11), 3350–3371,
- Mitchell, H. L., and P. L. Houtekamer, 2009: Ensemble Kalman Filter Configurations and Their Performance with the Logistic Map. *Monthly Weather Review*, **137** (12), 4325–4343,
- Mittermaier, M., 2019: A meta analysis of the Fractions Skill Score: The limiting case and implications for aggregation. *Monthly Weather Review*, MWR-D-18-0106.1,
- Mittermaier, M., and N. M. Roberts, 2010: Intercomparison of Spatial Forecast Verification Methods: Identifying Skillful Spatial Scales Using the Fractions Skill Score. *Weather and Forecasting*, **25** (1), 343–354,
- Mittermaier, M., N. M. Roberts, and S. A. Thompson, 2013: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorological Applications*, **20** (2), 176–186,
- Murphy, A. H., 1993: What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, **8** (2), 281–293,
- Nastrom, G. D., and K. S. Gage, 1985: A Climatology of Atmospheric Wavenumber Spectra of Wind and Temperature Observed by Commercial Aircraft. *Journal of the Atmospheric Sciences*, **42** (9), 950–960,
- Necker, T., M. Weissmann, and M. Sommer, 2018: The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quarterly Journal of the Royal Meteorological Society*, **144** (714), 1667–1680,
- Nielsen, E. R., and R. S. Schumacher, 2016: Using Convection-Allowing Ensembles to Understand the Predictability of an Extreme Rainfall Event. *Monthly Weather Review*, **144** (10), 3651–3676,
- Nuss, W. A., and D. K. Miller, 2001: Mesoscale predictability under various synoptic regimes. *Nonlinear processes in Geophysics*, **8** (6), 429–438,

- Palmer, T. N., 2000: Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics*, **63** (2), 71–116,
- Palmer, T. N., A. Dörnbrack, G. Seregin, A. Döring, and G. Seregin, 2014: The real butterfly effect. *Nonlinearity*, **27** (9), R123–R141,
- Panosetti, D., S. Böing, L. Schlemmer, and J. Schmidli, 2016: Idealized Large-Eddy and Convection-Resolving Simulations of Moist Convection over Mountainous Terrain. *Journal of the Atmospheric Sciences*, **73** (10), 4021–4041,
- Peralta, C., Z. Ben Bouallègue, S. E. Theis, C. Gebhardt, and M. Buchhold, 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *Journal of Geophysical Research Atmospheres*, **117** (7), 1–13,
- Picard, L., and C. Mass, 2017: The Sensitivity of Orographic Precipitation to Flow Direction: An Idealized Modeling Approach. *Journal of Hydrometeorology*, **18** (6), 1673–1688,
- Piper, D., M. Kunz, F. Ehmele, S. Mohr, B. Mühr, A. Kron, and J. Daniell, 2016: Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany Part 1: Meteorological background. *Natural Hazards and Earth System Sciences*, **16** (12), 2835–2850,
- Potvin, C. K., and L. J. Wicker, 2013: Assessing Ensemble Forecasts of Low-Level Supercell Rotation within an OSSE Framework. *Weather and Forecasting*, **28** (4), 940–960,
- Putnam, B. J., M. Xue, Y. Jung, G. Zhang, and F. Kong, 2017: Simulation of Polarimetric Radar Variables from 2013 CAPS Spring Experiment Storm-Scale Ensemble Forecasts and Evaluation of Microphysics Schemes. *Monthly Weather Review*, **145** (1), 49–73,
- Reinhardt, T., and A. Seifert, 2006: A three-category ice scheme for LMK. *COSMO Newsletter*,
- Richardson, L., 1922: Weather Prediction by Numerical Process.
- Ritter, B., and J.-F. Geleyn, 1992: A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations. *Monthly Weather Review*, **120** (2), 303–325,
- Roberts, N. M., and H. W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, **136** (1), 78–97,
- Robinson, F. J., S. C. Sherwood, and Y. Li, 2008: Resonant Response of Deep Convection to Surface Hot Spots. *Journal of the Atmospheric Sciences*, **65** (1), 276–286,
- Rodwell, M. J., and Coauthors, 2013: Characteristics of Occasional Poor Medium-Range Weather Forecasts for Europe. *Bulletin of the American Meteorological Society*, **94** (9), 1393–1405,
- Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing Forecast Error in a Convection-Permitting Ensemble System. *Monthly Weather Review*, **142** (12), 4519–4541,

- Rotunno, R., and C. Snyder, 2008: A Generalization of Lorenz's Model for the Predictability of Flows with Many Scales of Motion. *Journal of the Atmospheric Sciences*, **65** (3), 1063–1076,
- Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Periañez, and R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society*, **142** (696), 1453–1472,
- Schwartz, C. S., G. S. Romine, M. L. Weisman, R. A. Sobash, K. R. Fossell, K. W. Manning, and S. B. Trier, 2015: A Real-Time Convection-Allowing Ensemble Prediction System Initialized by Mesoscale Ensemble Kalman Filter Analyses. *Weather and Forecasting*, **30** (5), 1158–1181,
- Schwartz, C. S., and R. A. Sobash, 2017: Generating Probabilistic Forecasts from Convection-Allowing Ensembles Using Neighborhood Approaches: A Review and Recommendations. *Monthly Weather Review*, **145** (9), 3397–3418,
- Schwartz, C. S., and Coauthors, 2009: Next-Day Convection-Allowing WRF Model Guidance: A Second Look at 2-km versus 4-km Grid Spacing. *Monthly Weather Review*, **137** (10), 3351–3372,
- Schwartz, C. S., and Coauthors, 2010: Toward Improved Convection-Allowing Ensembles: Model Physics Sensitivities and Optimizing Probabilistic Guidance with Small Ensemble Membership. *Weather and Forecasting*, **25** (1), 263–280,
- Seifert, A., C. Köhler, and K. D. Beheng, 2012: Aerosol-cloud-precipitation effects over Germany as simulated by a convective-scale numerical weather prediction model. *Atmospheric Chemistry and Physics*, **12** (2), 709–725,
- Selz, T., and G. C. Craig, 2015: Upscale Error Growth in a High-Resolution Simulation of a Summertime Weather Event over Europe. *Monthly Weather Review*, **143** (3), 813–827,
- Skok, G., 2015: Analysis of Fraction Skill Score properties for a displaced rainband in a rectangular domain. *Meteorological Applications*, **22** (3), 477–484,
- Skok, G., 2016: Analysis of Fraction Skill Score properties for a displaced rainy grid point in a rectangular domain. *Atmospheric Research*, **169** (October 2014), 556–565,
- Skok, G., and N. M. Roberts, 2018: Estimating the displacement in precipitation forecasts using the Fractions Skill Score. *Quarterly Journal of the Royal Meteorological Society*, **144** (711), 414–425,
- Slingo, J., and T. N. Palmer, 2011: Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **369** (October), 4751–4767,
- Snook, N. A., M. Xue, and Y. Jung, 2011: Analysis of a Tornadoic Mesoscale Convective Vortex Based on Ensemble Kalman Filter Assimilation of CASA X-Band and WSR-88D Radar Data. *Monthly Weather Review*, **139** (11), 3446–3468,
- Snook, N. A., M. Xue, and Y. Jung, 2012: Ensemble Probabilistic Forecasts of a Tornadoic Mesoscale Convective System from Ensemble Kalman Filter Analyses Using WSR-88D and CASA Radar Data. *Monthly Weather Review*, **140** (7), 2126–2146,

- Snook, N. A., M. Xue, and Y. Jung, 2015: Multiscale EnKF Assimilation of Radar and Conventional Observations and Ensemble Forecasting for a Tornadoic Mesoscale Convective System. *Monthly Weather Review*, **143** (4), 1035–1057,
- Snyder, C., T. Bengtsson, P. Bickel, and J. L. Anderson, 2008: Obstacles to High-Dimensional Particle Filtering. *Monthly Weather Review*, **136** (12), 4629–4640,
- Snyder, C., and F. Zhang, 2003: Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter*. *Monthly Weather Review*, **131** (8), 1663–1677,
- Sobash, R. A., and D. J. Stensrud, 2013: The Impact of Covariance Localization for Radar Data on EnKF Analyses of a Developing MCS: Observing System Simulation Experiments. *Monthly Weather Review*, **141** (11), 3691–3709,
- Stephan, K., S. Klink, and C. Schraff, 2008: Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Quarterly Journal of the Royal Meteorological Society*, **134** (634), 1315–1326, arXiv:0801.1618v2.
- Sterk, A. E., D. B. Stephenson, M. P. Holland, and K. R. Mylne, 2016: On the predictability of extremes: Does the butterfly effect ever decrease? *Quarterly Journal of the Royal Meteorological Society*, **142** (694), 58–64,
- Stevens, S., C. Schreck, S. Saha, J. Bell, and K. Kunkel, 2019: Precipitation and fatal motor vehicle crashes: continental analysis with high-resolution radar data. *Bulletin of the American Meteorological Society*,
- Stratman, D. R., M. C. Coniglio, S. E. Koch, and M. Xue, 2013: Use of Multiple Verification Methods to Evaluate Forecasts of Convection from Hot- and Cold-Start Convection-Allowing Models. *Weather and Forecasting*, **28** (1), 119–138,
- Sun, Y. Q., R. Rotunno, and F. Zhang, 2017: Contributions of Moist Convection and Internal Gravity Waves to Building the Atmospheric 5/3 Kinetic Energy Spectra. *Journal of the Atmospheric Sciences*, **74** (1), 185–201,
- Sun, Y. Q., and F. Zhang, 2016: Intrinsic versus Practical Limits of Atmospheric Predictability and the Significance of the Butterfly Effect. *Journal of the Atmospheric Sciences*, **73** (3), 1419–1438,
- Supinie, T. A., N. Yussouf, Y. Jung, M. Xue, J. Cheng, and S. Wang, 2017: Comparison of the Analyses and Forecasts of a Tornadoic Supercell Storm from Assimilating Phased-Array Radar and WSR-88D Observations. *Weather and Forecasting*, **32** (4), 1379–1401,
- Surcel, M., I. Zawadzki, and M. K. Yau, 2014: On the Filtering Properties of Ensemble Averaging for Storm-Scale Precipitation Forecasts. *Monthly Weather Review*, **142** (3), 1093–1105,
- Surcel, M., I. Zawadzki, and M. K. Yau, 2015: A Study on the Scale Dependence of the Predictability of Precipitation Patterns. *Journal of the Atmospheric Sciences*, **72** (1), 216–235,
- Surcel, M., I. Zawadzki, M. K. Yau, M. Xue, and F. Kong, 2017: More on the scale dependence of the predictability of precipitation patterns: Extension to the 2009–13 CAPS spring experiment ensemble forecasts. *Monthly Weather Review*, **145** (9), 3625–3646,

- Thompson, P. D., 1957: Uncertainty of Initial State as a Factor in the Predictability of Large Scale Atmospheric Flow Patterns. *Tellus*, **9** (3), 275–295,
- Vié, B., O. Nuissier, and V. Ducrocq, 2011: Cloud-Resolving Ensemble Simulations of Mediterranean Heavy Precipitating Events: Uncertainty on Initial Conditions and Lateral Boundary Conditions. *Monthly Weather Review*, **139** (2), 403–423,
- Walser, A., D. Lüthi, and C. Schär, 2004: Predictability of Precipitation in a Cloud-Resolving Model. *Monthly Weather Review*, **132** (2), 560–577,
- Walser, A., and C. Schär, 2004: Convection-resolving precipitation forecasting and its predictability in Alpine river catchments. *Journal of Hydrology*, **288** (1–2), 57–73,
- Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2008: SALA Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review*, **136** (11), 4470–4487,
- Weston, P. P., W. Bell, and J. R. Eyre, 2014: Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, **140** (685), 2420–2429,
- Weyn, J. A., and D. R. Durran, 2017: The Dependence of the Predictability of Mesoscale Convective Systems on the Horizontal Scale and Amplitude of Initial Errors in Idealized Simulations. *Journal of the Atmospheric Sciences*, **74** (7), 2191–2210,
- Weyn, J. A., and D. R. Durran, 2018a: Ensemble Spread Grows More Rapidly in Higher-Resolution Simulations of Deep Convection. *Journal of the Atmospheric Sciences*, **75** (10), 3331–3345,
- Weyn, J. A., and D. R. Durran, 2018b: The scale dependence of initial-condition sensitivities in simulations of convective systems over the southeastern United States. *Quarterly Journal of the Royal Meteorological Society*,
- Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-Scale Data Assimilation and Ensemble Forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar Data Experiments. *Weather and Forecasting*, **30** (6), 1795–1817,
- Yang, S.-C., E. Kalnay, B. R. Hunt, and N. E. Bowler, 2009: Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter. *Quarterly Journal of the Royal Meteorological Society*, **135** (638), 251–262,
- Yano, J.-I., and Coauthors, 2018: Scientific Challenges of Convective-Scale Numerical Weather Prediction. *Bulletin of the American Meteorological Society*, **99** (4), 699–710,
- Yussouf, N., D. C. Dowell, L. J. Wicker, K. H. Knopfmeier, and D. M. Wheatley, 2015: Storm-Scale Data Assimilation and Ensemble Forecasts for the 27 April 2011 Severe Weather Outbreak in Alabama. *Monthly Weather Review*, **143** (8), 3044–3066,
- Yussouf, N., and D. J. Stensrud, 2010: Impact of Phased-Array Radar Observations over a Short Assimilation Period: Observing System Simulation Experiments Using an Ensemble Kalman Filter. *Monthly Weather Review*, **138** (2), 517–538,

- Zeng, Y., U. Blahak, and D. Jerger, 2016: An efficient modular volume-scanning radar forward operator for NWP models: description and coupling to the COSMO model. *Quarterly Journal of the Royal Meteorological Society*, **142** (701), 3234–3256,
- Zeng, Y., U. Blahak, M. Neuper, and D. Jerger, 2014: Radar beam tracing methods based on atmospheric refractive index. *Journal of Atmospheric and Oceanic Technology*, **31** (12), 2650–2670,
- Zeng, Y., T. Janjić, A. DeLozar, S. Rasp, U. Blahak, A. Seifert, and G. C. Craig, 2019: Comparison of methods accounting for subgrid-scale model error in convective-scale data assimilation.
- Zhang, F., N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale Predictability of Moist Baroclinic Waves: Convection-Permitting Experiments and Multistage Error Growth Dynamics. *Journal of the Atmospheric Sciences*, **64** (10), 3579–3594,
- Zhang, F., A. M. Odins, and J. W. Nielsen-Gammon, 2006: Mesoscale Predictability of an Extreme Warm-Season Precipitation Event. *Weather and Forecasting*, **21** (2), 149–166,
- Zhang, F., Y. Qiang Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019: What is the Predictability Limit of Midlatitude Weather? *Journal of the Atmospheric Sciences*, JAS–D–18–0269.1,
- Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of Moist Convection on Mesoscale Predictability. *Journal of the Atmospheric Sciences*, **60** (9), 1173–1185,
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. *Monthly Weather Review*, **132** (5), 1238–1253,
- Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2015: Practical Predictability of the 20 May 2013 Tornadic Thunderstorm Event in Oklahoma: Sensitivity to Synoptic Timing and Topographical Influence. *Monthly Weather Review*, **143** (8), 2973–2997,
- Zhang, Y., F. Zhang, D. J. Stensrud, and Z. Meng, 2016: Intrinsic Predictability of the 20 May 2013 Tornadic Thunderstorm Event in Oklahoma at Storm Scales. *Monthly Weather Review*, **144** (4), 1273–1298,
- Zimmer, M., G. C. Craig, C. Keil, and H. Wernli, 2011: Classification of precipitation events with a convective response timescale and their forecasting characteristics. *Geophysical Research Letters*, **38** (5), n/a–n/a,
- Zwibel, J., J. Van Baelen, S. Anquetin, Y. Pointin, and B. Boudevillain, 2016: Impacts of orography and rain intensity on rainfall structure. The case of the HyMeX IOP7a event. *Quarterly Journal of the Royal Meteorological Society*, **142**, 310–319,

Acknowledgements

First and foremost, I want to thank my supervisor George Craig, and Christian Keil and Martin Weissmann for giving helpful advice in all aspects of this project. They enabled me to make unique experiences in research and in life. Furthermore, I want to thank Bernhard Mayer for co-examining this thesis.

I am grateful to Tobias Selz for always being thought-provoking and helpful and for constantly pushing me to reach if not for the stars then certainly for the clouds—mostly by making me hike through them.

Thank you, Mirjam Hirt, for asking interesting and difficult questions and disagreeing about the answers, as well as sharing a profound fascination for the tiny details of a visualization. Also for telling me the truth even if it means I have to rewrite my conclusions.

I fell deep gratitude towards Yvonne Ruckstuhl for translating and mediating between math and me. And for all those driven but fruitless arguments about whether the best way to explain something is a plot or an equation.

I want to thank Florian Baur for always taking some time to discuss a finding or just the weather and for being a friend and a regular support throughout the minor troubles of being a PhD student.

Thanks to Robert Redl and Leonhard Scheck for hours of helping out with computer or coding issues, life advice, and so much more.

And thanks to Stephan Rasp for making me a better climber and, therefore, colleague.

I also want to thank my office mates, Matthias Schindler and Tobias Necker, for keeping up with me, generating a pleasant and productive working atmosphere and for showing me that not everything at work is about work.

Finally, I'd like to thank my parents for supporting me without actually knowing what I was doing all those years. Over the last years, my amazing partner Julia has been the most patient, adventurous and loving person, making every hardship seem easy.

Quotes

Perhaps some day in the dim future it will be possible to advance the computations faster than the weather advances and at a cost less than the saving to mankind due to the information gained. But that is a dream.

- Lewis Fry Richardson

In an earlier paper dealing with predictability, the writer quoted a meteorologist, whose identity he still cannot recall, as having maintained somewhat disparagingly that if the theory of atmospheric instability were correct, one flap of a sea gull's wings would forever change the future course of the weather. If we take the results of the present study at face value, we might conclude in addition that such a change would be realized within about seventeen days.

- Ed Lorenz

Chaos: When the present determines the future, but the approximate present does not approximately determine the future.

- Ed. Lorenz

Prediction is very difficult, especially about the future.

- Karl Kristian Steincke

Truth is much too complicated to allow anything but approximations.

- John von Neumann

Young man, in mathematics you don't understand things. You just get used to them.

- John von Neumann

Since all models are wrong, the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

- George Box

The days are over, of hanging out the seaweed, examining the size of molehills, or studying animal entrails for portents of coming tempests - that is, unless the computers are down!

- Tim Palmer

Predictability is to prediction as romance is to sex!

- 'A famous climatologist', published by Tim Palmer

The good thing about science is that it's true whether or not you believe in it.

- Neil deGrasse Tyson